FINAL REPORT

COOPERATIVE AGREEMENT CR 810239

THE ESTIMATION OF RECREATION-RELATED WATER POLLUTION
CONTROL BENEFITS: SWIMMING, BOATING, AND
MARINE RECREATIONAL FISHING

by

William J. Vaughan
Charles M. Paulsen
Julie A. Hewitt
and
Clifford S. Russell

Submitted by Resources for the Future

to

U.S. Environmental Protection Agency

August 1985

DISCLAIMER

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

ESTIMATING WATER POLLUTION CONTROL BENEFITS USING
PARTICIPATION MODELS: EXECUTIVE SUMMARY

The research summarized in this report was conducted for the U.S. Environmental Protection Agency in support of that agency's efforts to comply with the requirements of President Reagan's Executive Order 12291 of February 1981. That order made cost-benefit analysis of proposed major regulatory decisions mandatory, even when that analysis could not, by law, be the basis for actual decisions. Where environmental regulations are in question, the techniques and databases necessary to successful cost-benefit analysis are still far from sufficiently developed to support routine applications.[1] Accordingly, the Office of Policy Analysis within EPA has supported research with the dual goal of developing data and methods and of generating actual benefit estimates.

The specific charge for this study was to develop methods for estimating the benefits of water pollution control as they accrue to society through effects on participation of individuals in swimming (in natural water bodies), recreational boating, and recreational fishing in Great Lakes and marine water of the United States. This particular combination of activities was chosen as representing the major promising extensions to the work done in an earlier project on the benefits accruing via freshwater recreational fishing. (Vaughan and Russell, 1982).

To put the effort in perspective, it is worth reviewing the benefit estimates collected and critically reviewed by Freeman (1982). In table 9.1 (p. 170) Freeman gives as his "most likely point estimate" for total annual water pollution control benefits (1978 dollars) $9.4 billion.

Almost half this total ($4.6 billion) he estimates in chapter 8 to be accounted for by "Recreation". This broad category he in turn has broken down into five subcategories: freshwater fishing, marine sports fishing, boating, swimming and waterfowl hunting. The last category accounts for a tiny percentage (about 2 percent) of the $4.6 billion total. Swimming and the two types of fishing are estimated to account for about 22 percent each (about $1.0 billion per year), while boating is estimated to account for almost 33 percent (or $1.5 billion). Thus, if successful, a project providing improved benefit estimates and methods for future estimation exercises for fishing, boating, and swimming would enhance EPA's ability to deal with a substantial fraction of the currently estimated total benefits to be expected from the ongoing U.S. program of water pollution control.

We believe that overall the project has been successful. As will be documented below, substantial progress has been made in clarifying the promise and problems inherent in the traditional participation method for dealing with recreation decisions. Further, a very different method for producing benefit estimates has been adapted to the peculiarities of the available data on boating participation. The application of the traditional methods to the fishing and swimming activity categories was, however, only partially successful. The lesson we draw from partial success is, however, of some importance in its own right, for it points to some fundamental data gaps that must be filled before benefit estimation in these areas can be really successful, let alone a routine operation for a regulatory agency.

The remainder of this chapter is devoted to a general introduction to conceptual problems and opportunities with the participation method, and to an anticipatory summary of major results. This will set the stage for

chapters dealing with each of the conceptual problems in more detail
(2,3,4); and chapters describing applications to fishing (5,6,7), and
swimming (8.9). In chapters 10 and 11, a different technique is explored
in estimating benefits accruing via boating. Here the focus is on
ownership decisions (of the necessary capital good, the boat) rather than
on activity decisions about days of boating. This method has much to
recommend it where the data are available to support it. A concluding
chapter pulls together the important conclusions and re-emphasizes the
implications for future research.

PROBLEMS WITH PARTICIPATION ANALYSIS AS THE BASIS FOR BENEFIT ESTIMATION

For more than two decades, applied economic models of consumer
recreation decisions have proceeded along two parallel tracks, the
parallelism seemingly dictated by the nature of the data available for
model estimation (Cicchetti, Fisher and Smith 1973).

The "macro" track has been characterized by the recreation
participation equation approach. It involves an attempt to estimate a
relationship explaining the pattern and intensity of individual
participation in specific recreational activities at a national or regional
level of spatial analysis, regardless of the places (sites) where the
activities took place. The "micro" track, in contrast, is characterized by
travel cost models attempting to econometrically capture the demand
relationship for the services of a single known site or group of sites.

While both the "macro" and "micro" approaches appear to stand on firm
theoretical foundations, those foundations were not build independently of
the way data on recreation happens to be collected. The data for
estimating the "macro" relations comes from large cross-sectional

population surveys containing information on the socio-economic characteristics of the respondents, leisure activities in which they engaged (if any), and their intensity of participation in these activities over a specified time span (for example, see U.S.D.I n.d.). Such population surveys generally contain minimal information on where the respondents recreated or could have recreated. Consequently, analysts have no information on the vector of individual-specific travel costs indexed by activity category and site (which are analogous to goods prices in models of consumer demand for marketed goods as discussed by Wennergren 1967) associated with enjoying the several activity categories surveyed. Thus, elaborate theoretical reasoning has been brought to bear simply to explain. why such price-type data is not needed in model estimation.

This lack of precise individual-specific price data, along with a focus on activities rather than sites, and the collection of information on non-participants as well as participants are the features distinguishing the "macro" approach from the site-specific travel cost method (Cicchetti, Fisher and Smith 1973). They are also the features which, unfortunately, have confounded our understanding of just what the macro model represents - a structural demand equation or a reduced form - and have rendered welfare analysis with it extremely tenuous.

The study reported here focussed on three questions central to the application of the macro or participation techniques to the estimation of benefits of a policy change. First, what is the place in the models of measures of availability or quality of the relevant resource (fishable water to participation in fishing). Such measures are necessary to the reflection of the results of policy and hence to prediction of post-policy. participation, which are in turn the basis for benefit estimates. It is

therefore not surprising that measures of availability <u>do</u> appear in participation. What they are doing there is another matter.

Second, it is necessary to look carefully at how participation equations are actually estimated given the special nature of the data available on consumer choices. Thus, in the last decade or so, estimation of microeconomic models of consumer behavior using large individual- or household-level data sets has flourished and proven an important advance in applied economics. Details typically masked in aggregate time-series data analysis are often available in individual cross-sectional data, thus allowing the testing of hypotheses about responses of individual or household demand-supply bundle choices to changes in constraints.

It is in such micro datasets that one tends to find measures of demands and supplies that economists would characterize either as corner solution realizations of instantaneous optimizing decisions or as discrete representations of such decisions. An example of the former case would be where one has data on the number of hours an individual worked in the market over a given year, and for some subset of individuals no market hours were worked. An instance of the latter case, is where data are available only on whether or not an individual had purchased some consumer durable over the previous twelve months, but not on the amount of the expenditure. Assuming the statistical models determining labor supply and durables demand to be the objectives of estimation, then the former is an example of what have come to be known as limited dependent variable (LDV) models, while the latter is a member of the class of qualitative dependent variable (QDV) models. Tobin's pioneering 1958 paper on durables demand is the forerunner of LDV estimation in economics. Using data on 735 micro spending units, Tobin modeled the ratio of durables expenditures to

disposable income; for 183 of these spending units, no durables were purchased during the time period of interest and a "corner solution" had to be treated. The solution to this problem was the genesis of the Tobit estimator, which will be discussed below in chapter 4. Note that had Tobin only data on whether or not there was some durable purchased rather than on the actual amount, a QDV model (such as binary probit or logit) would have been the appropriate approach.

In recreation participation modeling, owing to the nature of the available micro data, standard econometric techniques such as ordinary least squares (OLS) will typically be inappropriate tools for the analysis of recreation participation. The available data on participation decisions, rather, are usually qualitative or limited dependent variables, and more complicated estimation techniques are in general required in order to obtain consistent estimates of the parameters governing the participation outcomes. Maximum likelihood is the estimation method most commonly used in such analysis.

In previous empirical analyses of recreation participation (Vaughan and Russell (1982)), efforts were focused less on the subtleties of the statistical and econometric methods used than on the development of a unified framework for assessing empirically the effects of water quality changes on participation. As such, the analysis was restricted to those econometric methods that were less resource-consuming than is true of many of the more sophisticated iterative maximum likelihood techniques to be described below in chapter 4. This strategy was followed consciously, though at the expense of the possible inconsistencies resulting from ignoring such subtleties. In the present endeavor, emphasis is shifted to an evaluation of the implications of certain characteristics of the

participation measures (e.g., nonnegativity of some measures, discrete or grouped measures) for estimation strategy and, as a consequence, for benefit estimation.

The third question to ask of the participation method is, How, if at all, can we obtain dollar values from it. The "micro" travel cost model, being a structural representation of a single demand function or a demand system can be employed directly to produce site values. It can also be used to assess the welfare change occasioned by adding or deleting a site from a pre-specified system of sites, or to answer other welfare-related questions, such as the benefits of upgrading site quality, as well. Its primary limitation is an arbitrary definition of the scope of the problem, specifically the identification of a subsystem of sites which can reasonably be modeled without omitting relevant substitutes.

The participation equation "macro" model begs this question by ignoring sites per se. But, since prices do not appear as independent variables in the model specification - due primarily to data deficiencies in our view - direct welfare analysis with such models would seem to be impossible. But the macro models are used, in an indirect way, for welfare analysis. Indeed, their primary practical purpose has been the prediction of changes in participation levels over time or across space under alternative hypothetical public policies directed toward the supply of recreational resources. These changes, and hence the policies engendering them, are usually valued using a unit value which Freeman 1982 graciously refers to as an "activity shadow price". The monetary welfare measures assigned to the possible policy alternatives are then obtained as the product of the predicted change in days of participation, summed over the population, and the unit value. (See Freeman 1982 for a catalog of several

such studies, including one by the authors of this paper. Researchers with the perspicacity _not_ to assign a dollar value to projected participation changes include Hay and McConnell 1979.)

To those familiar with the site-specific travel cost approach, the unit day value method may seem no more than an irrelevant curiosum, but in fact its use is commonplace. It is a practice that was recommended, until recently, by the Water Resources Council and was cited recently as an alternative when other methods were not available (WRC 1979). It has been used to value an entire recreational fishery in British Columbia (Pearce Bowden 1971), and to estimate the recreational benefits of the Illinois river in Oklahoma under the Wild and Scenic rivers Act (U.S. Department of the Interior 1979). Other agencies, including the Corps of Engineers, continue to use this method. Moreover, the Forest Service, in responding to requirements of the multiple use and sustained yield legislation has incorporated the equivalent of unit day values in their programming models (Sorg, _et_. _al_., 1984). Finally, the method has found frequent application in analyses of the national recreation benefits of water pollution control programs, as catalogued by Freeman 1982.

The unit day value approach to obtaining the welfare effects of a policy change is particularly convenient when no information other than a prediction of a policy's impact on days of participation is available from a macro participation model. But, there are three problems with the "macro" modeling approach: differential site quality characteristics are not accounted for; prices are often omitted in estimation; and unit values are employed to monetize predicted quantity changes. The first two problems lead to biased predictions of quantity change while the first and third distort the estimate of welfare change, even if the quantity

predictions are accurate.   Only rarely are these limitations acknowledged (an exception is Sorg, et. al., 1984).

AN EXPERIMENT WITH A MORE DIRECT METHOD: THE CASE OF RECREATIONAL BOATING

The conventional participation equation method of estimating the recreational benefits of water quality improvement policies focuses on the changes the probability and intensity of leisure activity participation occasioned by the prospective policies.   Since, for many such activities, lumpy expenditures for activity-specific consumer durables are not a prerequisite for participation, their role is ignored in both theoretical model development and econometric estimation.   Instead, trip cost or a proxy thereto plays a paramount role   (See chapter 2 and Vaughan and Russell, 1984).

There are two compelling reasons why the standard participation equation method may not be the best approach to the investigation of the benefits of water quality improvement accruing to activities such as recreational boating.   One is theoretical and one purely data-related. First, boating obviously requires a boat, which can either be rented or owned.   It also requires an environmental service, boatable quality water. So, from the theoretical side, these obvious relationships suggest that Mäler's notion of weak complementarity (Mäler 1974, pp. 131-139) between an indivisible private good (the boat durable) and a public good (boatable quality water), can usefully be employed.   The notion implies that a portion of benefits of a change in water quality can be identified via the estimation of the demand relation for the private good, the boat durable, which itself depends on the level of provision of the public good. This is the portion of the total benefit accruing to new entrants to boat

recreation, assuming standard operating year participation levels. It is estimated in chapter 10. The remaining portion of the benefits, that which accrues to existing boat owners who participate more intensively, is estimated in chapter 11.

In this connection the consumer can be regarded as having a direct utility function defined over discrete durable goods, a numeraire composite commodity, and the service flows from the durable, if it is owned. In the boating case, the service flow can be proxied by the number of boating trips enjoyed over the year. We choose to model this problem in two steps (for a more sophisticated treatment see Hanemann 1984). In the first step the discrete ownership probabilities for various boat types are econometrically modeled via conditional logit. In the second step the demand for the continuous quantity, boat trips, conditional on ownership, is modeled via regression analysis, under the assumption that the conditional density (conditional in the sense that boat ownership is chosen) of the positive realizations is log normal. The procedure is analogous to one variant of the class of hurdles models discussed in Cragg 1971 and chapter 4 below. It departs from the unified treatment in Hanemann 1984 in the sense that the estimating equation for the continuous choice is not derived from an underlying theoretical utility maximization model, although the discrete choice model is. Rather, the continuous quantity equation estimated in our second step is best regarded as an approximation. The practical significance of this shortcut is that an overall compensating variation measure of welfare change associated with more boat ownership and more trips due to pollution control cannot be obtained. Rather, while a compensating variation measure is available for the first benefit component, we must settle for an ordinary Marshallian

surplus for the second.   (For an example of how to get a theoretically consistent   compensating   variation   welfare   measure   out   of   a discrete/continuous utility model once all utility function parameters are statistically estimated see Hanemann 1982a.)

But, before becoming involved with the details of discrete choice modeling, another important consideration should be raised.   This is the nature of the available data.   As noted above, participation-type models require detailed information on individuals, both in terms of the choices they made and their characteristics.   Unfortunately, the best source of data on boating, the 1976 Coast Guard Nationwide Boating Survey, contains almost none of the information required by the participation equation approach.   Particularly,   the absence of good socioeconomic data on respondents,   (especially non-owners) rules out the application of the participation equation method to the Coast Guard Survey data.   But, the survey's detailed information on boat ownership by type of boat and category of recreational use, number of boat outings per household, and trailer miles per outing, make it an attractive source of data which, when supplemented by independent information on boat costs, can be put to use in estimating a model of discrete (durable goods) demand, and a continuous model of trip demand conditional on boat ownership.   Thus, both theory and practical necessity drive toward this experiment.

ANTICIPATING THE RESULTS

The results produced by the several subprojects briefly described above are of interest principally because they challenge what has become the conventional wisdom in this area, the numbers pulled together and critiqued by Freeman 1982. anticipating the summary and comparisons to be

provided in the final chapter of this report, we can say that every benefit number reported here is lower than the corresponding number in Freeman's key table, after the latter are adjusted to 1983 prices by the Consumer Price Index. Indeed, all but one of our numbers -- the lower limit of our Great Lakes and marine fishing benefits -- are lower in 1983 terms than Freeman's 1978 dollar figures.

How much lower? for all the activities dealt with in this report, quite substantially. comparing most likely point estimates, we find the highest relative RFF number to be 30 percent of Freeman's Great Lakes and marine recreational fishing. The lowest, for boating and swimming together, is only five percent of Freeman's corresponding figure. (See table 1.1).

One must, of course, treat <u>all</u> the numbers in these comparisons with great caution. Freeman has outlined the problems in the earlier studies he has synthesized. The chapters of this report will emphasize the difficulties and uncertainties that plague our work. It is worth saying for a first time here, however, that the major problem is with data. Most fundamentally, it is impossible to find comprehensive water quality data for the pre-pollution-control situation of the basis of which participation or other relations can be estimated. The data that are available are not comprehensive in geographic coverage, consistent in quality and coverage of particular pollutants, and for the most part do not include parameters that can reasonably be hypothesized to enter into recreationist's views of availability of water resources for their activity decisions. For example, systematic work on swimming benefits requires data on the extent of pollution <u>relevant to</u> swimming, such as turbidity and micro organism counts. These pieces must be available comprehensively enough that we can

Table 1.1.   RFF and Freeman Benefit Estimates Compared
(109 dollars, 1983, per year)

	RFF		Freeman[f]	RFF as % of Freeman
Freshwater Fishing	0.9^a		1.5	60
Marine & Great Lakes Fishing	0.4^b		1.5	30
Boating	0.2^c	NA	2.3	9
Swimming	0^d	NA	1.5	nmf
Boating and Swimming	0.2	0.01^e	3.8	(5)(0)
Totals	1.5	1.3	6.8	(39(19)

n.m.f. = no meaningful figure.

Notes:

a.   From Vaughan and Russell 1982 with correction to 1983 using the CPI.

b.   From chapter 7 below.

c.   From chapter 11 below; using the complementary good method.

d.   From chapter 9 below; a generous interpretation of the results for swimming alone in table 9.26.

e.   From chapter 9 below; the mean of the overall benefit estimates in table 9.26.

f.   From Freeman 1982.

characterize the reduction in availability of swimmable water due to pollution at the level of the state or, preferably, the county. They are not now available in anything like that detail.

The estimates produced in this report depend not on such comprehensive objective data, but rather on the largely subjective characterizations reported to us by responsible state officials. While these are at least reasonably comprehensive (See appendix C to chapter 5) the nature of the survey process that produced them, must give us pause in their interpretation.

Still, these survey numbers are arguably more to the point than the characterizations lying behind the earlier studies, and it seems reasonable to think that conventional wisdom about the likely size of participation-based benefits is due for readjustment. This is not to say anything about the so-called intrinsic beneits being sought in other EPA-supported projects.

PLAN OF THE BOOK

The reader whose major or only interest is in the benefit measures just briefly discussed is invited to skip ahead at this point to chapter 12, the final chapter, where the numbers are summarized. Those readers who have some interest in methodological issues per se will want to begin at the beginning, with chapters 2 through 4. These cover, respectively the role of recreation resource availability variables in participation analysis; the pitfalls of two-step (probability/intensity) estimation of participation - based benefits; and some of the more recent developments in estimation techniques for models using qualitative, truncated or censored dependent variables.

Subsequent to these general methodological discussions we have three major sections in which the specifics of data and method are presented for marine and Great Lakes recreational fishing, swimming (in natural water bodies but not disaggregated by type of water) and recreational boating (also not divided into fresh and saltwater) respectively. The chapters involved are:

fishing: 5-7

swimming: 8,9

boating: 10,11

The final chapter, as indicated above, provides a summary of the benefit estimates.

NOTES

1.   This discussion abstracts from conceptual problems of doing <u>any</u> cost benefit analysis for certain kinds of proposed regulations.   For example, where the effects of the regulation apply specifically to the discharge of one industry.   The benefits can only be determined by making arbitrary assumptions about what is happening to other industries.   This is because individual plants occur in multi-industry regions rather than in single-industry clumps.

REFERENCES

Cicchetti, Charles J., Anthony C. Fisher, and V. Kerry Smith. 1973. "Economic Models and Planning Outdoor Recreation," Operations Research, vol. 21, no. 5 (September-October) pp. 1104-1113.

Freeman, A Myrick III. 1982. Air and Water Pollution Control: A Benefit-Cost Assessment. (New York: John Wiley and Sons), 186 pp.

Hanemann, W. Michael. 1982. "Quality and Demand Analysis," in Gordon C. Rausser, ed. New Directions in Econometric Modeling and Forecasting in U.S. Agriculture (New York: Elsevier North Holland) pp. 55-98.

Hanemann, W. Michael. 1984. "Discrete/Continuous Models of Consumer Demand," Econometrica, vol. 52, no. 3 (May), pp. 541-561.

Hay, Michael J. and Kenneth E. McConnell. 1979. "An Analysis of Participation in Nonconsumptive Wildlife Recreation," Land Economics, vol. 55 (November) pp. 460-471.

Maler, Karl-Goran. 1974. Environmental Economics. (Baltimore, Md.: Johns Hopkins University Press for Resources for the Future).

Pearce Bowden Economic Consultants Limited. 1971. "The Value of Fresh Water Sport Fishing in British Columbia," (Victoria, British Columbia: Department of Recreation and Conservation).

Sorg, Cindy, John Loomis, Dennis M. Donnelly and George Peterson. 1984. "The Net Economic Value of Cold and Warm Water Fishing in Idaho," Unpublished Working Paper, (Collins, Colorado: U.S. Forest Service).

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables," Econometrica, vol. 26, pp. 24-36.

U.S. Department of the Interior. n.d. 1975 National Survey of Hunting, Fishing and Wildlife-Associated Recreation (NSHFWR) Data tapes and unpublished mimeo, state and national reports (Washington, D.C.).

Vaughan, William J., and Clifford S. Russell. 1982. "Valuing a Fishing Day: an Application of a Systematic Varying Parameter Model," Land Economics, vol. 58, no. 4 (November), pp. 450-463.

Vaughan, William J., and Clifford S. Russell. "The Role of Recreation Resource Availability Variables in Recreation Participation Analysis," Journal of Environmental Management, vol. 19, (December).

Water Resource Council. 1979. "Procedures for Evaluation of National Economic Development (NED) Benefits and Costs in Water Resources Planning (Level C), Final Rule," Federal Register, vol. 44, no. 242, December 14, 1979, pp. 72892-977.

Wennergren, W. Boyd. 1967. "Surrogate Pricing of Outdoor Recreation," Land Economics, vol. 43, no. 1 (February), pp. 112-116.

Chapter 2

THE ROLE OF RECREATION RESOURCE AVAILABILITY VARIABLES IN PARTICIPATION ANALYSIS*

Suppose a decision on providing or not providing some general addition to recreation resources hinges on what impact the addition is projected to have on participation in the activities to which they are relevant. For example, suppose a decision about expanding camping areas across the U.S. is to be made on the basis of the projected addition to camping activity attributable to the addition of resources. This problem setting allows us to postpone until later consideration of the problems of valuation within the participation model context.

To address the problem a cross-sectional data set reflecting individual leisure-time pursuits and the socio-economic characteristics of the same individuals is required, so that population leisure participation can be estimated econometrically as a function of these characteristics, as in Settle (1980). It also seems necessary to have variables measuring the supply of recreation resources appear as arguments in the equations to be estimated, so that the effect of alterations in supply can be appraised directly. But, a question arises at this point: Do such supply variables belong in recreation participation equations, in the sense that the equation specification is consistent with economic theory?

A hint of the answer is given by the travel-destination/modal-choice literature, where relevant independent variables in the empirical model of choice are the variables that would appear in the consumer's indirect utility function--for example travel cost (analogous to goods prices), site

attributes, consumer income and consumer characteristics (Hensher and Johnson 1981, Rugg 1973, Small and Rosen 1981). Unfortunately, few, if any, recreation participation surveys from a broad sample of the population contain detailed individual-specific information on travel and other costs incurred in going from place of residence to the recreation site or sites chosen, let alone other potential sites not chosen. Nor do the surveys normally identify the location of individuals or sites at all precisely. Thus, if a correctly specified recreation participation equation is to be estimated econometrically from such survey data, a proxy variable must be developed which can stand in, however crudely, for the expected site prices associated with an individual's participation in one or more recreational activities. Fortunately, this variable is indeed a resource supply variable.

Previous empirical analyses of population recreation participation in broad activity categories (rather than site-specific travel cost studies) have either employed a measure of average variable travel cost consistent with theory (Ziemer and Musser 1979; Ziemer, et. al., 1982) or, when such measures were unavailable from survey data, substituted aggregate "supply" variables as proxies (Davidson, Adams and Seneca 1966; Chicchetti 1973; Deyak and Smith 1978; Smith and Munley 1978; Hay and McConnell 1979; Vaughan and Russell 1982) or even ignored the problem entirely (Settle 1980). The rationale for such proxy recreation resource supply variables has generally been vaguely asserted rather than clearly established. Yet it makes intuitive sense to link participation to the "availability" of recreation alternatives measured in terms of quantity (number of facilities in a geographic region) or quality (number of facilities per capita to account for congestion) (Cicchetti, Fisher and Smith 1973). In fact, it is

possible to go beyond intuition and provide a firm rationale for the inclusion of explanatory physical supply quantity variables in recreation participation equations. We do so below, using the case of a water-based recreation activity (eg. fishing).

A version of the theory of distance estimators of density (or in our case density estimators of distance) developed in the statistical ecology literature can be applied to show that expected travel cost should be functionally related to the number of water bodies per unit land area in a region.

RELATING DENSITY AND DISTANCE

The idea behind this link is intuitively appealing, the more objects there are randomly arranged in a given space, the closer will be the nearest such object on average, to any randomly chosen point. If we knew the parameters of the process that put the objects in their places, we could obtain an exact expression for the expected distance. However, we will usually not know either the exact process behind the location or the parameter appropriate to an approximate process. In those circumstances, which characterize the analyst looking at actual water bodies in regions and wondering about a proxy for travel cost, observed density of the bodies may be used either directly or after transformation as a proxy for expected distance.

To tie down the intuitive idea with a bit more rigor, assume that a region can be divided into N equal-size squares. These squares will be taken to be units. Some number, n of "tiles" representing water bodies and also of unit size, will be placed on the grid by a random process such that the probability of a "tile" falling on a square is 1/N=P. More than one

tile can land on a square, so that after all tiles have been placed, the observed number of "lakes" will be $w \leq n$. If N is large (p small) the resulting probabilities of a particular number, m, of tiles falling on any chosen grid square can be approximated by the Poisson density function:

$$P(m;np) = \frac{e^{-np}(np)^m}{m!}$$

The expected number of water bodies, allowing for multiple tiles per square, is $N(1-e^{-np}) = w$. Because e^{-np} can be approximated by the first few terms of the series

$$1 - np + \frac{np^2}{2!} - \frac{np^3}{3!} + \cdots$$

and because $np = n/N < 1$ by assumption, it is also true that w/N, the observed density of lakes, is an approximation for np, the Poisson parameter (often written as λ).

Thus, $w/N = 1-e^{-np} = 1-(1-np) = np = n/N$

This approximation result is important when the objects on a grid may be assumed to have been distributed according to a Poisson density function with parameter λ. Then it is possible to show that the expected distance $E(r)$ from a randomly chosen point to the nearest such object is given by

$$E(r) = \int_0^\infty r 2\pi r \lambda e^{-\lambda \pi r^2} dr$$

$$= \int_0^\infty 2\pi \lambda r^2 e^{-\lambda \pi r^2} dr$$

The derivation of this expected distance formula is reasonably straightforward. By the Poisson distribution the probability of no objects in a circle of radius r is:

$$P(\lambda \pi r^2 = 0) = e^{-\lambda \pi r^2}$$

If the nearest object appears at distance r from the center of this circle,

we can define an annular ring of width dr within which it is the only such

object.   The area of the annular ring is

$$\pi(r + dr)^2 - \pi r^2 = \pi(r^2 + 2rdr + dr^2 - r^2)$$

$$= \pi(2rdr + dr^2)$$

Ignoring terms in $(dr)^2$ we can approximate the probability that the band

contains the one object by

$$e^{-\lambda 2\pi r\ dr}(2\pi r\lambda dr)^1/1\ !$$

using the reasoning developed above.   Note, however, that

$$xe^{-x} \cong x(1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \ .\ .\ .)$$

$$\cong x - x^2 + \frac{x^3}{2!} - \frac{x^4}{3!} + \ .\ .\ .$$

Since $x = \lambda 2\pi rdr$, and ignoring terms of order 2 and higher in dr, we have

$$2\pi rdr\ \lambda e^{-2\pi rdr\lambda} \cong 2\pi rdr\lambda$$

If the two events (no objects within the area πr^2; one object within

the annular ring with area $\pi(2rdr + dr^2)$) are assumed to be independent

their joint probability is the product of their individual probabilities.

Thus the joint probability density function of distance r is the product of

the Poisson probability expressions for finding zero objects out to r and 1

object in the narrow band at r

$$f(r) = 2\pi r\lambda e^{-\lambda\pi r^2}dr$$

Thus, the expected value of r, or of the average distance to the nearest

object from random points in the space, as a function of the density

parameter is:

$$E(r) = \int_0^\infty r \, 2\pi r \lambda e^{-\lambda \pi r^2} \, dr = \int_0^\infty 2\pi \lambda r^2 e^{-\lambda \pi r^2} \, dr$$

This definite integral can be shown to produce:

$$E(r) = \frac{1}{2} \lambda^{-1/2}$$

which is to say that the expected distance from a randomly chosen point to the nearest object depends on the Poisson parameter. Thus, if we can approximate λ by w/N, we can approximate E(r) by $1/2(w/N)^{-1/2}$ so that expected distance falls with increasing density: This relation is shown in figure 2.1.

The variance in expected distance (VAR r) can be obtained by recognizing (Larsen and Marx 1981, p. 114) that VAR(r) equals $E(r^2)$ − $(E(r))^2$. The expected value of r is already known to be $0.5\lambda^{-0.5}$ so the second term in VAR(r) is this quantity squared, equal to $0.25\lambda^{-1}$. To obtain the expected value of r^2, we take the definite integral:

$$E(r^2) = 2\pi \lambda \int_0^\infty r^3 e^{(-\lambda \pi r^2)} \, dr$$

$$= 2\pi \lambda \, \frac{\Gamma(2)}{2(\pi \lambda)^2} = \frac{2\pi \lambda}{2(\pi \lambda)^2} = \frac{1}{\lambda \pi}$$

So,

$$\text{Var}(r) = \frac{1}{\pi \lambda} - \frac{1}{4\lambda}$$

Putting this expression in terms of a common denominator and simplifying

$$\text{VAR}(r) = \frac{(4-\pi)}{4\pi \lambda} \cong 0.068\lambda^{-1}.$$
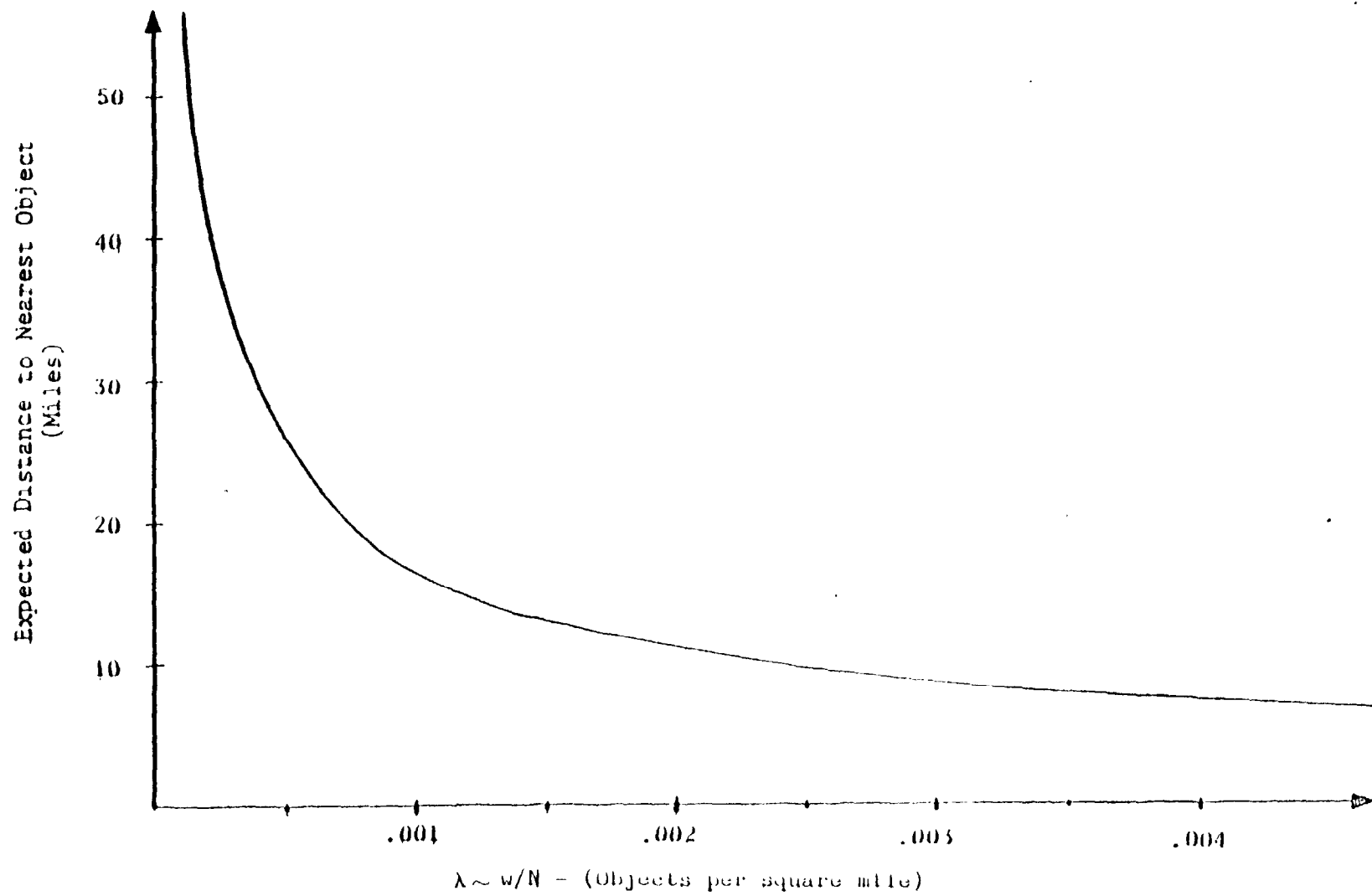
Figure 2-1. The Density Distance Relationship

and the variance of the expected distance also falls with density.

While these relations have both intuitive appeal and formal justification, there are several possible pitfalls associated with using a measure of the density of water bodies (acres per acre) in the region of interest as an inverse proxy for distance and hence travel cost.

First, the relation between measured density as a point estimate of expected density and λ is better the smaller λ. This may be seen by inspecting the series approximation for e^{-np} given above. The smaller n relative to N the more rapidly the terms with exponents greater than one approach zero. Thus, the more richly endowed the region the less reliable the approximation.

Second, in the real world water bodies do not come as discrete unit area pieces, or indeed as pieces of any common size across a single region let alone across several regions. Thus, the assumptions underlying the derivation will be violated in actual regions. Particularly, data on surface acreage (rather than the number of lakes) is the most common measure of the availability of water for recreation, and surface acreage is composed of lakes of varying sizes as well as rivers and streams. So the Poisson forest analogy does not translate perfectly in application.

To see the problem let $\hat{\lambda}$, measured as the square miles covered by the objects (lakes) per square mile of regional surface area, be the available data. Suppose that all objects have the same size, m, so that λ (number of units) = $\hat{\lambda}/m$. Then,

$$E(r) = 0.5\lambda^{-0.5} = 0.5\,(\hat{\lambda}/m)^{-0.5} = 0.5\hat{\lambda}^{-0.5}(m^{0.5})$$

So, if m is constant across regions, λ can be used as a proxy for λ as an explanatory variable in estimating activity participation relationships, since the constant term $(m^{0.5})$ will merely scale the estimated availability parameter.   A plausible assumption is that large lakes are composed of clusters of equal radii objects, so proportionality between λ and λ is maintained.   It is however, implausible to think that m will be constant across regions; and finding a set of region specific average m_s is neither practically non-theoretically appealing.

Third, even if the objects of interest are of uniform size across the regions, but their locations were generated by a heterogenous, nonrandom process rather than a homogenous Poisson process (i.e., the objects' centers were not uniformly and independently distributed) the expected distance formula will not hold (Ripley 1981, Ch 7, 8).

Finally, if the intensity parameter varies from place to place but the manner in which it varies is unknown a priori, spatial groupings cannot be established which uniquely reflect the variation in the several population λ's associated with the different regions.   All one can do is to produce different area-weighted mean density proxy measures for λ for different levels of aggregation across space.

For example, in a 100 by 100 grid, we generated two samples with 400 objects (λ = .04) and two samples with 200 objects (λ = .02).   The distance to the nearest object was computed from 81 points systematically located at the intersection of lines of latitude and longitude ten units apart. (Border intersections were excluded).   The expected value of distance to the nearest object is 2.5 miles for λ = .04 and 3.54 miles for λ =.02. The sample outcomes for expected distance and the associated standard errors of the means from this simple experiment show that in these cases the sample

means are all within one standard error of the population expectation given
by $0.5\lambda^{-0.5}$:

	λ = .02		λ = .04	
	Sample 1	Sample 2	Sample 1	Sample 2
Sample Mean Distance	3.38	3.61	2.41	2.59
Std Error of Mean	0.19	0.20	0.16	0.13
Theoretically Expected Distance	3.54		2.50	

Note, however, that if we were to sample over both grids believing
that both belonged to the same population (i.e., shared the same λ) our
estimate of λ would be (200 + 400)/2(10,000) or 0.03 and our expected
distance would be 2.89. Although this expected distance would perhaps be
realistic for individuals located on or around the border delineating the
regions (particularly the geographic centroid of the two regions together)
it would not be for individuals located some distance from that border, who
more properly should be assigned their respective region - specific
expected distances.

SOME IMPLICATIONS FOR AGGREGATION: MEASURING THE PROXY FOR λ

With aggregate real world data we do not pick a set of random points in
space and mark off the distance from each of those points to the closest
"object" (i.e., water body), to estimate a value for λ from the inverse of
expected distance formula. Rather we use acres of objects per acre of
total area as a proxy for λ and hence for expected distance. The question
is how to demarcate the relevant boundaries of total regional areas?
Counties, combinations of counties, or fixed areas around each individual

could be used, but the cutoff distance over which our proxy for λ should be measured is unknown.[1]

However, a University of Kentucky Water Resources Institute survey (Bianchi, 1969) of over 3,000 fisherman reported that only slightly more than 8 percent travelled over 30 miles to fish. Similar calculations of the percent of days fishing by travel distance can be made from U.S. Department of the Interior, 1982:

One-Way Distance (miles)	Frequency (%)
0-5	19
6-24	25
25-49	17
50-99	14
100-249	10
250-499	3
500-999	1
>1000	Nil

The median travel distance from this data is 32 miles. The Davies test of skewness (Langley 1970) suggests this data is approximately logarithmic in distribution, so the geometric mean is appropriate, yielding a value of 31.6 miles. It also appears that 250 miles would be a generous upper limit for the radius of the region whose characteristics determine recreationist behavior. Two alternatives, then, suggest themselves. One is to use density data only from an individual's county of residence. At the other extreme, circular regions around the centroid of the individual's county of residence could be constructed and weighted density data from all the counties represented in this region used to construct a measure of λ.

CONCLUSION

It is appropriate to include two "availability" variables in the econometric analysis of recreation participation choice; one to capture the distance or travel cost influence via the number (or acres) of recreational resource facilities per unit land area and one to capture the (expected) congestion influence via the number (or acres) of such facilities per capita.

Further, it is reasonable to maintain that individuals base their recreation participation decisions on expected (travel-cost based) prices across the gamut of alternative types of recreation activities rather than actual prices, since the latter cannot always be known with certainty for a broad array of activities. In this case availability variables are hot just proxies introducing errors-in-variables problems into the econometric analysis (Maddala 1977, Ch. 13). Rather, these observed variables are the true price variables which we desire to measure based on the theoretical model. In this context errors-in-variables problems would occur only if the degree of spatial aggregation involved in constructing a measure of λ was too coarse, encompassing several areas which belonged to separate populations, each with its own particular λ. In such a situation it is likely that the estimated parameter reflecting the relationship between participation and average availability will be a biased measure of the true effect.

NOTES

1.   For further discussion and numerical illustrations relevant to this aggregation problem, see Vaughan. et. al., 1985.

REFERENCES

Beyer, William H., ED.   1981.   CRC Standard Mathematical Tables 26th ed.
     (Boca Raton: CRC Press) p. 342.

Bianchi, Dennis H.   1969.   The Economic Value of Streams for Fishing,
     Research Report No. 25,   (University of Kentucky: Water Resources
     Institute).

Cicchetti Charles J.   1973.   Forecasting Recreation in the United States
     (Lexington, Mass.: Lexington Books).

Cicchetti, Charles J., Anthony C. Fisher, and V. Kerry Smith.   1973.
     "Economic Models and Planning Outdoor Recreation," Operations Research
     vol. 21, no. 5 (September-October) pp. 1104-1113.

Davidson, Paul, F., Gerald Adams, and Joseph Seneca. 1966. "The Social
     Value of Water Recreational Facilities Resulting from an Improvement
     in Water Quality: The Delaware Estuary," in Allen V. Kneese and
     Stephen C. Smith, eds., "Water Research (Baltimore, Md.:   Johns Hopkins
     University Press for Resources for the Future).

Deyak, Timothy A. and V. Kerry Smith. 1978.   "Congestion and Participation
     in Outdoor Recreation: A Household Production Function Approach,"
     Journal of Environmental Economics and Management vol. 5, no. 1
     (March) pp. 63-80.

Hay, Michael J. and Kenneth E. McConnell.   1979.   "An Analysis if
     Participation in Nonconsumptive Wildlife Recreation," Land Economics
     vol. 55 (November) pp. 460-471.

Hensher, David, A. and Lester W. Johnson.   1981.   Applied Discrete-Choice
     Modelling.   (New York: John Wiley and Sons.) Ch. 2, 3.

Langley, Russell.   1970.   Practical Statistics Simply Explained.   (New
     York: Dover) Ch. 3.

Larsen, Richard J. and Morris L. Marx.   1981.   An Introduction to
     Mathematical Statistics and Its Applications.   (Englewood-Cliffs:
     Prentice Hall).   pp. 536.

Maddala, G.S.   1977.   Econometrics (New York: McGraw-Hill) Ch. 13, pp.
     292-300.

Pollard, J. H.   1971.   "On Distance Estimators of Density in Randomly
     Distributed Forests." Biometrics, vol. 17 (December), pp. 991-1002.

Pollard, J. H.   1977.   A Handbook of Numerical and Statistical Techniques.
     (London: Cambridge University Press) pp. 110-111.

Ripley, Brian, D.   1984.   Spatial Statistics.   (New York: John Wiley and
     Sons) Ch. 7. pp. 130-34.

Rugg, Donald. 1973.   The Choice of Journey Destination: A Theoretical and Empirical Analysis," Review of Economics and Statistics, vol. 15, no. 1, (February), pp. 64-72.

Settle, J. G. 1980.   "Relating Participation in Recreational Activities to Social Characteristics," Journal of Environmental Management vol. 10, pp. 138-147.

Small, Kenneth A. and Harvey S. Rosen.   1981.   "Applied Welfare Economics with Discrete Choice Models," Econometrica, vol. 49, no. 1 (January), pp. 105-130.

Smith, V. Kerry and Vincent G. Munley. 1978.   "The Relative Performance of Various Estimators of Recreation Participation Equations," Journal of Leisure Research vol. 10, no. 3, pp. 167-176.

Unwin, David. 1981.   Introductory Spatial Analysis.   (New York: Methuen Inc.) Chapter 3, pp. 29-62.

U.S. Department of the Interior. Fish and Wildlife Service. 1982. 1980 National Survey of Fishing, Hunting and Wildlife - Associated Recreation. (Washington, D.C.: GPO).

Vaughan, William J. and Clifford S. Russell.   1982.   Freshwater Recreational Fishing, the National Benefits of Water Pollution Control.   Washington, D.C.: Resources for the Future).

Vaughan, William J., John Mullahy, Julie A. Hewitt, Michael Hazilla and Clifford S. Russell, 1985.   "Aggregation Problems in Benefit Estimation: A Simulation Approach," unpublished report to U.S. EPA from Resources for the Future, Washington, D.C.   (August).

Ziemer, Rod F. and Wesley N. Musser.   1979.   "Population-Specific Recreation Demand Models and the Consequences of Pooling Sample Data," Western Journal of Agricultural Economics vol. 4 (July) pp. 121-129.

Ziemer, Rod, F., Wesley N. Musser, Fred C. White and R. Carter Hill. 1982. "Sample Selection Bias in Analysis of Consumer Choice: An Application to the Warmwater Fishing Demand," Water Resources Research, vol. 18, no. 2 (April) pp. 215-219.

Chapter 3

TWO STEP ESTIMATION OF PARTICIPATION BENEFITS

This chapter focuses on two problems with bifurcating the estimation
of the benefits of recreational resource enhancement into two unrelated
steps - quantity change and valuation.   Throughout we assume the absence of
systematic error in predicting participation change, though such errors can
either offset or compound the error attributable to assigning an average
unit value to that change.   The discussion is confined to macro
participation models of the aggregate level of recreation activity service
flow enjoyed at an (unknown) site or set of sites, rather than travel cost
models of the demand for the services of a site or system of sites, because
in the latter case the ability to estimate demand functions obviates the
need for unit values.

After a brief review of the genesis of the two-step method the
valuation problem and the theoretical background engendering it is
addressed.   Subsequently it is shown that the two-step valuation method is
questionable on theoretical grounds and not likely to provide a reasonably
accurate monetary measure of the welfare change associated with
participation quantity change stimulated by a policy of recreation resource
augmentation.   Some numerical examples are provided which confirm the
theoretical results.

ORIGINS OF THE TWO-STEP METHOD OF VALUATION

Traditionally, data on the regionally differentiated availability of
recreational resources (acreage or number of lakes, campgrounds, natural
forests, etc., contained in some geographic region, often the state) has

been obtained to supplement the data in population recreation surveys and included among the set of relevant regressors in the macro participation model specification. Inclusion of these variables has some basis in common sense, as it is intuitively appealing to anticipate that recreation resource availability variables must have sane role to play in influencing recreation participation and intensity. One would expect individuals living in a region amply endowed with freshwater to be more likely to engage in water sports, and do them more often over the year, ceteris paribus, than individuals living in a poorly endowed region. But more important, inclusion of such availability regressors in the model is <u>absolutely necessary</u> if it is to be a useful tool for evaluation of potential broad policies of supply alteration. If there are no supply variables in the participation equation, participation will not be predicted to change when supply changes, and the policy will appear to have no impact the analyst can value.

Initially, the inclusion of quantity-type availability variables in macro models was theoretically justified by somewnat vague allusions to "supply" factors (Cicchetti 1973), although it was never clear what sort of a supply function was implied. Later, to help dispel the confusion, Deyak and Smith (1978) invoked household production theory to explain supply in terms of the household's marginal cost for "producing" recreational service flows. Marginal cost itself for these authors was a function of "characteristics" variables, which happened to be physical availability variables in disguise, represented as facilities per capita, a measure of expected congestion. The consequence of this paradigm was the essential endogeneity of (self-supplied) price, which practically speaking meant that only reduced forms could conveniently be estimated. Since the household's

internally determined shadow price is never observed, that left only a reduced form quantity equation to be estimated. Hence the requirement of a second valuation step for welfare analysis.

But, as just demonstrated in chapter 2, the elaborate theoretical household production model is really not necessary as a vehicle to justify the inclusion of a measure of the quantity of recreational resources per unit land area in econometric models of recreation participation. This argument offers an explicit justification for using physical availability regressors as proxies for "average" travel-cost based activity prices in the direct estimation of an aggregate structural activity demand equation, rather than a reduced form. However, because of the expected-value nature of the proxy, parameter bias is the penalty imposed by using it in lieu of the correct individual-specific activity prices (McFadden and Reid 1975).

Yet the problem of placing a unit dollar value on the participation change occasioned by a particular policy of recreation resource enhancement to produce a monetary benefit measure remains. It is equally difficult, whether we believe we have estimated a reduced form activity participation. equation as a function of individual characteristics and site characteristics measured by some availability measure, or a structural activity quasi-demand equation with availability as a proxy for activity price. In neither case do we observe individual prices directly, and the best that can be done is to predict a quantity change conditional on a hypothesized change in availability, and value it arbitrarily in a second step.

To see why this valuation procedure gained currency, it is necessary to explore the theoretical background giving rise to a "participation" equation.

THEORETICAL BACKGROUND: IS THE PARTICIPATION QUANTITY EQUATION REALLY A REDUCED FORM?

In order to exploit the calculus, conventional utility theory makes the implicit assumption that the consumer's optimal consumption bundle will represent an interior solution in the space of available alternatives. That is, the maximum of the consumer's utility function occurs at an interior point of the budget space where all goods are consumed in positive amounts, not at a corner where one or more commodities are not consumed at all (Russell and Wilkinson, 1979).

Quandt (1970) observed that this implicit assumption is unrealistic in travel-oriented applications, since consumers do not "undertake a little bit of travel by every mode on every link in a network" (p. 5). The same observation can be made about leisure activities, since population surveys often reveal large proportions of non-participants. Thus, the implicit interior solution assumption of conventional utility theory must be relaxed, or the theory itself reformulated, in order to incorporate the phenomenon of non-participation (i.e., zero consumption).

The first alternative is to remain within the confines of traditional utility theory, relaxing the interior solution assumption. The corner solution rationale for zero consumption in leisure pursuits is made by Ziemer, et. al., (1982) based on the Kunn-Tucker conditions. Essentially this means ruling out the class of utility functions where the marginal rates of substitution between pairs of goods are everywhere defined and equatable to the respective goods price ratios. For example, members of the Bergson family of utility functions which are all transformations of the additive (in logarithms) homothetic utility function $U = \prod_{i=1}^{n} x_i$ are

ruled out, since their indifference curves never cut the goods axes, and corner solutions cannot occur.

Another route to explaining the same phenomena is to reformulate neoclassical utility theory along household production lines. In this "new" approach, the household does not obtain utility directly from purchased goods or recreation site visits. Rather, it employs these goods, along with its own time, to produce outputs of utility - yielding entities (non-market goods, service flows, wants, or characteristics depending on the author) over which the utility function is defined. (Cicchetti and Smith 1973, 1976).

There are two general variants of the household model - the Becker (1965) version and the Lancaster (1966) version, reviewed lucidly in Cicchetti and Smith (1973). The Lancaster version, utilized to analyze recreation choice by Rugg (1973), Mak and Moncur (1980), and Greig (1983) is particularly appealing because its general form guarantees zero consumption of some goods, independent of the class of utility function specified. Conventional utility theory can be regarded as a special case of the general Lancaster model where the production technology matrix is diagonal. In this latter instance, corner solutions can be produced by an appropriate formulation of the utility function. Therefore, the flexibility of the Lancaster model to represent <u>either</u> the neoclassical case with corner solutions <u>or</u> the "pure" Lancaster case makes it an obvious choice. But, either theoretical household production model yields roughly equivalent equations to be estimated from survey data explaining recreational trips. Particularly, the inclusion of income, site characteristics, and trip expenses or a physical resource availability proxy thereto is commonplace (See McConnell and Strand 1981, Rugg 1973, and

Ziemer, <u>et</u>. <u>al</u>., 1982 for superficially comparable "trips" equations

derived from different theoretical models).

　　　To exploit the calculus suppose the popular Becker household

production framework as outlined in Deyak and Smith 1978 is used, with the

restrictive assumption of non-jointness in production. If the individual's

recreation service flow production function is classically well behaved and

exhibits constant returns to scale, then the self-supply equation is

defined by the marginal cost, mc, of producing service flow q, and is

constant and independent of the levels of production of non-recreational

service flows. If trip cost is unobserved but is known to be a function,

t(a), of availability (Vaughan and Russell 1984) we get (1a.) below. If

site characteristics affect marginal household production cost via expected

congestion (Deyak and Smith 1978) represented by h(a) we get (1.b):

　　　mc = t(a)   (1.a)

　　　mc = $\partial c/\partial q$ = f(p,w,h(a))   (1.b)

where

　　　mc = individual's marginal cost

　　　 c = total cost

　　　 q = recreation service flow in constant quality units

　　　 p = prices of market goods

　　　 a = resource availability

　　　 w = wage rate

　　t(a) = trip costs as a function of availability.

　　h(a) = expected congestion as a function of availability.

　　　Each individual's inverse demand function can be expressed in terms

of income and tastes.　　If WP is the individual's marginal willingness to

pay for the service flows, and the utility function is of the Bergson family with zero cross-price effects:

$$WP = \frac{\partial TWP}{\partial q} = g\ (y,s,q) \tag{2}$$

where

y = individual's income

s = individual's tastes

TWP = total willingness to pay

The superstructure of (1.b) and (2) above can be recast into the neo-classical mold of (1.a) and (2) if additional restrictive assumptions are imposed (the zero cross-price assumption can be relaxed). Particularly, goods prices can be treated as exogenous if we assume:

- a fixed total leisure time constraint so there is no income - leisure tradeoff,

- trips of constant duration with zero fixed costs so the "price" of a trip is equivalent to its variable (travel) cost,

- activity categories within which sites are of fairly homogenous quality -eg. trout fishing in coldwater streams,

- a factor of proportionality converting site visits (or trips) in an activity category into a service flow uniquely related to that activity over which utility is defined.

In the Deyak and Smith model of Eq. 1.b and 2 individual equilibrium is determined where the level of q equates mc to WP. For estimation, the Deyak and Smith model produces two reduced form equations from the structural equations in (1.b) and (2), and estimates (3) below.

$$q = (y,\ s,\ p,\ h(a),\ w) \tag{3}$$

$$mc = WP = \phi(y,s,p,h(a),w) \tag{4}$$

In contrast, the neoclassical structural activity demand model estimates the demand function version of (2) directly. Since travel cost-based price is viewed as exogenous by equation (1.a), reduced form equations (3) and (4) are not required. The neoclassical model leads to a specification expressing q as a function of resource availability (or price, if available), income, and tastes.

An overriding consideration in all of this is the desire to arrive at an empirical specification which does not require price regressors, since such information is unavailable in most population recreation surveys.[1] Particularly, those who view availability variables as proxies for the quality of the experience, manipulate the household production model either to produce a reduced form quantity equation as above (if service flow outputs can be measured) or derive input demand equations (where site visits are treated as inputs). Irrespective of this sort of definitional legerdemain, in many applications the equation specification does not include price regressors, which happens to fit nicely with the character of the data. But, this practice is inconsistent with the theoretic& household model, be it a reduced form like Eq. 3 or, alternatively, a derived site visit input demand function.[2] (See Bockstael and McConnell 1983 for the latter, perhaps more reasonable, theoretical interpretation). Specification error bias is the obvious penalty paid for ignoring prices in this context. In contrast, those who argue that availability variables represent proxies for expected activity price attempt to define activity categories finely, so that quality is roughly constant, and view the estimated quantity equation as an activity demand curve. This specification also fits well with the survey data, but involves no inconsistency between theory and practice, because the estimated proxy

price parameter ideally should capture the true parameter, up to an unknown factor of proportionality. Because the availability proxy for the price represents expected travel distance, aggregation bias due to averaging over individuals (McFadden and Reid 1975) is a possible shortcoming.

VALUATION ISSUES

The conceptually correct Marshallian measure of benefits arising from increased resource availability may be written in terms of structural activity service flow supply and demand equations. Consider any individual, whose marginal cost of obtaining the recreation experience is a function of recreation resource availability:

mc^0 = marginal cost at pre-policy availability

a^0 particular to the individual,

mc^1 = marginal cost at post-policy availability

a^1 particular to the individual.

Suppose a policy of supply augmentation so $a^1 > a^0$. The individual's marginal willingness to pay for the experience is the demand price, WP, a function of the service flow quantity q. For the jth individual the net benefit of a policy of supply augmentation, $NB_j(a^0, a^1)$ can be written as:

$$NB_j(a^0, a^1) = \int_{q^0}^{q^1} WP(q)dq + q^0[mc^0 - mc^1] - mc^1[q^1 - q^0] \tag{5}$$

This expression is depicted graphically for two individuals in figure 3.1. The aggregate net benefit of the policy is the sum over all j=1,····· individuals of the net benefits in (5):

$$NB = \sum_{j=1}^{J} NB_j(a^0, a^1) \tag{6}$$
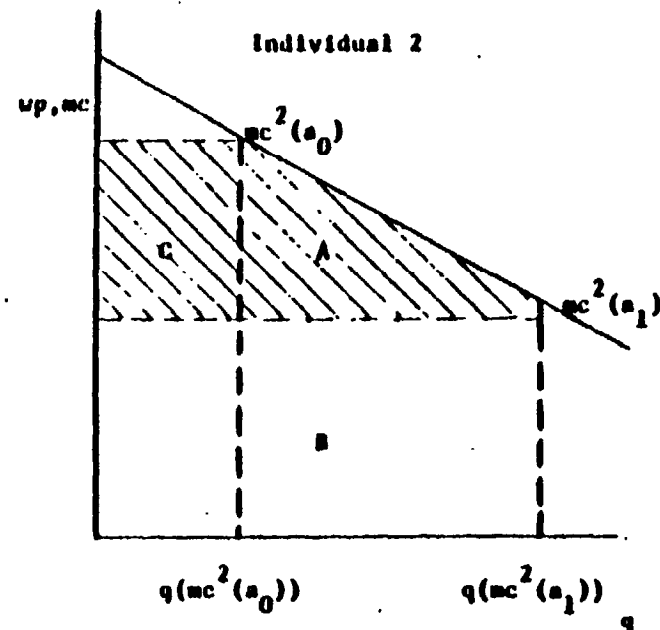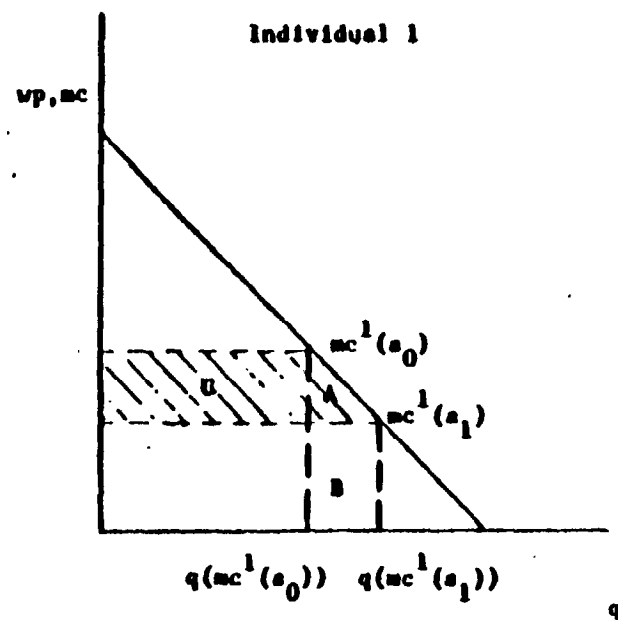
Individual 1

Individual 2

vp, mc

$mc^1(a_0)$

$mc^1(a_1)$

$mc^2(a_0)$

$mc^2(a_1)$

$q(mc^1(a_0))$ $q(mc^1(a_1))$

$q(mc^2(a_0))$ $q(mc^2(a_1))$

q

Legend

- $nb^1(a_0, a_1) = C + A = A + B + C - B = \text{\\\|}$

- $\displaystyle \int_{q(mc^1(a_0))}^{q(mc^1(a_1))} vp(q)\,dq = A + B$

- $(q(mc^1(a_0)))(mc^1(a_0) - mc^1(a_1)) = C$

- $\left(mc^1(a_1)\right)\left(q(mc^1(a_1)) - q(mc^1(a_0))\right) = B$

Figure 3.1. Individual net benefits of increased availability

As noted above, however, it would be very unusual to have the detailed individual data necessary to perform the calculation written so easily. A common situation is to have data allowing a prediction of the total increase in quantity produced from a macro participation model, $\sum_{j}(q_j^1 - q_j^0)$, and, from an independent source, a unit value to assign to the quantity change.

For instance Cicchetti, Fisher and Smith (1973) suggest:

> ..., the reduced-form participation equation can also be used, as we have suggested, to derive a measure of the benefits from a new facility. The amount of participation in an activity is first forecast under changing conditions of supply, i.e., without and then with the new facility. Then a measure of value or willingness to pay must be imputed to each unit (recreation day) of additional participation. Such measures have in the past been set for federal projects by water-resource agencies and approved by the U.S. Senate. Aggregate benefits are given by multiplying the imputed value per unit of participation by the change in the level of participation occasioned by the new facility. (p. 1011).

But, no explicit distinction is made by these authors between unit values which are conceptually equivalent to marginal willingness to pay (i.e., activity prices or, in the household model, unobserved shadow prices) and unit values which instead represent average willingness to pay over all units consumed (i.e., average consumer's surplus for the activity), although they seem to have had the former in mind.

A survey of the unit value literature reveals that most reported values are approximations to average, not marginal, willingness to pay (Dwyer, Kelly and Bowes 1977). If so, the direction of the valuation bias can be derived, and we do so below. But first, if we assume marginal unit values are available, can the procedure be justified?

Valuation with Marginal Unit Values

As McKenzie 1983 observes, there are two routes to welfare measurement; the familiar one where consumer demand functions are known, allowing direct computation of the Marshallian surplus measure; and alternative index-number approximations based on "only the prices and quantities that hold in alternative situations but not information about the shape of preferences or the consumer demand functions" (p. 101). The two-step valuation method in this context is a particularly simplistic version of this second route.

While the adjective marginal may evoke a subconsciously sympathetic response, valuation of a quantity change with marginal unit values (prices) does not guarantee a close approximation to the Marshallian consumer's surplus measure of welfare change, let alone the desired measures the latter approximates, compensating and equivalent variation. To demonstrate, begin with the most general case where a single price changes. Although the consumer's demand function for the good whose price has changed is unknown, assume that the quantity changes for all goods in the consumer's choice set are known, as are the initial and final price vectors.

When a single price changes the product of the n dimensional row vector of all n goods prices (measured at either their pre-policy levels, p^0, post policy levels, p', or an average of the two) and the n dimensional column vector of quantity changes can be used to produce welfare approximations if the demand function for the good whose price has changed is unknown (McKenzie 1983, Ch. 6.; Deaton and Muellbauer 1980, Ch. 7). These measures are known respectively as the Laspeyres and Paasche quantity

variation indices (LQV, PQV), and Harberger's consumer surplus (HCS).
Representing the marginal utility of expenditure in situation j as λ^j and
the utility index as U:

$$LQV = \sum p_i^0 \Delta q_i \approx \Delta U / \lambda^0 \tag{7}$$

$$PQV = \sum p_i^1 \Delta q_i \approx \Delta U / \lambda^1 \tag{8}$$

$$HCS = 1/2(LQV+PQV) \approx 1/2(\Delta U / \lambda^0 + \Delta U / \lambda^1) \tag{9}$$

where i = 1,..., N goods.

It can be shown that the HCS measure is an approximation to
Marshallian consumer surplus, since it simply takes the short-cut of
assuming the Marshallian demand curve is linear in the region of the price
change (Deaton and Muellbauer 1980, p. 188; McKenzie 1983, pp. 109-111).

However, the two-step valuation method, lacking information on the
own-good demand function and the quantity changes taking place outside the
market of direct interest, is more restrictive than the general case
represented by (7), (8) and (9). It deals more narrowly with the product
of price and quantity change for the good whose price has changed, ignoring
quantity changes for all other goods. So, the partial measures analogous
to (4), (5) and (6), indexing the good whose price changes as i are:

$$\widetilde{LQV} = p_i^0 \Delta q_i \neq \Delta U / \lambda^0 \tag{10}$$

$$\widetilde{PQV} = p_i^1 \Delta q_i \neq \Delta U / \lambda^1 \tag{11}$$

$$\widetilde{HCS} = 1/2(\widetilde{LQV} + \widetilde{PQV}) \neq 1/2(\Delta U / \lambda^0 + \Delta U / \lambda^1) \tag{12}$$

The partial index number measures assume, perhaps incorrectly, zero
cross-price effects. Except for unusually restrictive demand systems (eg:
Cobb-Douglas) when the price of a single good, i, changes, the quantities
of some other goods $j \neq i$ will change as well. But if other goods quantities
change, the partial LQV, PQV and HCS measures used in recreation benefits
analysis which ignore the sum of $p_j \Delta q_j$ for all $i \neq j$ are unlikely to bring us

reasonably close to the ideal welfare change measures, compensating variation (CV) and equivalent variation (EV), or even to the approximation they bound, Marshallian consumers surplus (CS).

The only case where quantity changes in other goods induced by a change in the price of the jth good can be ignored in calculating $\tilde{LQV}$, $\tilde{PQV}$ and $\tilde{HCS}$ is when the elasticity of demand for the jth good is unitary in absolute value over the region of interest. To prove this, arrange the arc price elasticity of demand formula (where e represents the absolute value of the arc elasticity) as:

$$1/2(p_i^1+p_i^0)(p_i^1-q_i^0)) \equiv 1/2(q_i^1+q_i^0)(p_i^0-p_i^1 e \qquad (13)$$

The l.h.s. of (13) is the definition of the partial Harberger consumer surplus measure, $\tilde{HCS}$. Expansion of the r.h.s. reveals that it represents the arc elasticity measure, e, times an approximation to the Marshallian consumer surplus integral $\widetilde{CS}$ obtained by linearizing the (unknown) demand curve between q_i^0 and q_i^1

$$\widetilde{CS} = 1/2(q_i^1+q_i^0)(p_i^0-p_i^1) = q_i^0(p_i^0-p_i^1) + 1/2(q_i^1-q_i^0)(p_i^0-p_i^1) \qquad (14)$$

The two expressions following the second equality in (14) represent the familiar welfare rectangle and triangle measures of Marshallian surplus.

So, the l.h.s. of (13) representing the partial measure $\tilde{HCS}$ either understates, overstates, or equals the approximate Marshallian consumer surplus measure on the r.h.s. depending upon whether the absolute value of the arc price elasticity of demand for the good whose price has changed is respectively less than, equal to, or greater than one.

But from (13) and (14), there is obviously no reason to compute the partial Harberger surplus measure $\widetilde{HCS}$ if q_i^0, q_i^1, p_i^1 and p_i^1 are all known or if q_i^0, p_i^1, q_i^1 and e are known, permitting calculation of p_i^1. In these circumstances the approximation $\widetilde{CS}$ can be obtained directly by linearizing the unknown demand function between p_i^0, q_i^0 and p_i^0, q_i^1 and applying (14). Of course the more nonlinear the demand function and the larger the price change the poorer the quality of the approximation $\widetilde{CS}$ to the correct measure CS. More often, only p_i^0 and the quantity change are known and no assumption is made about e; the welfare change being approximated instead by $\widetilde{LQV}$. Only under unusual circumstances will $\widetilde{LQV}$ equal $\widetilde{CS}$ defined in (14). For instance, assume the unknown demand function is of the linear form q=a-bp. Substitute this relation for the q_i^0 and q_i^1 terms in (15) defining the ratio of $\widetilde{CS}$ to $\widetilde{LQV}$ to get (12):

$$\widetilde{CS}/\widetilde{LQV} = 1/2(q_i^1+q_i^0)(p_i^0-p_i^1)/p_i^0(q_i^1-q_i^0) \tag{15}$$

$$\widetilde{CS}/\widetilde{LQV} = (2a-b(p_i^1+p_i^0))/2p_i^0 b \tag{16}$$

Since p_i^0 is exogenously given the function $(1-\widetilde{CS}/\widetilde{LQV})$ can be minimized with respect to p_i^1. The value $\overset{*}{p_i^1}$ which sets (16) equal to one is:

$$\overset{*}{p_i^1} = (2a-3bp_i^0)/b \tag{17}$$

From (17) if the initial evaluation point p_i^0, q_i^0 happens to be at the point of unit elasticity of the unknown demand function so $p_i^0 = 1/2 \ a/b$, substitution into (17) reveals $\overset{*}{p_i^1} = p_i^0 = 1/2 \ a/b$. The practical relevance of this first result is that if, by fortuitous accident, the initial point

of evaluation is at or very close to unit elasticity and the price change is small, the measure $\tilde{LQV}$ may not diverge overmuch from $\tilde{CS}$, but will deteriorate as p_i^1 becomes increasingly distant from p_i^0. Second (17) suggests that even if p_i^0 is not at the point of unit elasticity, there is a p_i^1 (and by implication a value of q_i^1 which sets (16) equal to 1. But there is no guarantee that the $\overset{*}{p_i^1}, \overset{*}{q_i^1}$ combination from (17) will be in the economic region ($\overset{*}{p_i}$ could be negative) or, if it is, that the policy being evaluated will throw up the $\overset{*}{p_i^1}, \overset{*}{q_i^1}$ combination that justifies the use of $\tilde{LQV}$. Finally, when price changes are "small", $\tilde{LQV}$, $\tilde{PQV}$ and $\tilde{HCS}$ will be approximately equal, but, unless the underlying unknown demand function is unit elastic over the region of change, none of them will be good approximations to $\tilde{CS}$.

In conclusion, it normally will not be possible to even compute the full LQV, PQV or HCS measures in the participation equation version of recreation benefits analysis, because changes in the consumer's entire consumption bundle remain unquantified. Without knowledge of the demand function, partial measures are unlikely to be representative of even a crude Marshallian consumers surplus measure of individual welfare changes, unless the utility function is such that unitary elasticity demand functions result (eg: Cobb Douglas) or the price change happens to be in the unit elastic region of an arbitrary demand function. While these conditions salvage the $\tilde{HCS}$ measure, if they are not met it is uncertain whether the sum of the unadjusted $\tilde{HCS}$ measures across all individuals will or will not be a useful aggregate. But, can anything be salvaged by using an average willingness to pay unit value rather than a marginal one? The answer, unfortunately, is not encouraging.

Valuation with Average Unit Values

In the usual case, only a measure $\overline{CS}$ of individual j's average surplus for the quantity of recreation activity (usually dollars per day) undertaken before a price change in activity i is available. In terms of the demand expression (2) it may be written as:

$$\overline{CS} = \int_{0}^{q^{0}} WP(q)dq/q^{0} \tag{18}$$

Under what circumstances, then, is the following approximation for net benefits a good one?

$$\widetilde{NB} = \overline{CS}(q^{1}-q^{0}) \tag{19}$$

We previously examined this question theoretically, using a representative consumer's situation, for the simplest inverse demand function, a linear one, $p = a + bq$, and for a constant elasticity function with elasticity n, $p = A^{1/\eta}q^{-1/\eta}$. If the ratio $q(a^{1})/q(a^{1})$ is written as k, the following expressions were obtained (see Vaughan and Russell 1982 for a full derivation):[3]

$$\text{linear demand: } \frac{\widetilde{NB}}{NB} = \frac{1}{1+k} < \frac{1}{2} \tag{20}$$

$$\text{constant elasticity demand } \frac{\widetilde{NB}}{NB} = \left[1 - nq^{0}{}^{(\frac{1}{\eta} - 1)}\right]\left[\frac{1-k}{1-k^{(1-1/\eta)}}\right] \tag{21}$$

where NB is a Marshallian consumers surplus. In addition, if the demand function is of the semi-logarithmic form $q = \exp(a+bp)$, CS evaluated at q^{0} is $-(q^{0}/b)$, where $b < 0$. This yields an average surplus $\overline{CS}$ of 1/b. It can easily be shown that $\widetilde{NB}$, the product of this average surplus and a quantity

change given as $\exp(a)$ $(\exp(bp^1) - \exp(bp^0))$ is exactly equivalent to the Marshallian consumers surplus measure NB from the definite integral of $\int_{p^0}^{p^1} \exp(a + bp)dp$.

Thus, if individual's demand functions are all linear (or nearly linear in the relevant range) the application of an average surplus always understates the total Marshallian CS measure of the welfare change by a factor of at least 0.5. If the demand function is of the constant elasticity sort, the approximation can either be correct, understate, or overstate the individual's surplus. Only when the demand function is semi-logarithmic does the procedure produce the correct result. So, applying an average unit value to an aggregate quantity change is also dangerous, with unknown risks a positive or negative valuation bias, depending on the nature of the (unknown) demand function.

These theoretical results may seem bloodless and unconvincing. So in the next section some numerical examples are constructed which verify them and give a concrete idea of just how wrong the approximation can be.

SOME NUMERICAL EXAMPLES OF THE VALUATION PROBLEM

This section begins with some simple numerical examples which assume demand specifications with zero cross-price effects -- a constant unitary elasticity specification $q=100/p$ and linear specification $q=25-p$.[4] We demonstrate the workings of the marginal and average welfare measures discussed previously, and contrast their accuracy vis-a-vis the correct Marshallian measure, CS. All of these results are easily calculated by the reader and, by assumption, since no other goods quantities change in response to the price/quantity change of interest, all partial index number welfare measure are exactly equivalent to their full counterparts (ie: LQV

$= L\tilde{Q}V$, $PQV = \widetilde{PQV}$, and $HCS = \widetilde{HCS}$).

Next, a more complex example based on an arbitrary parameterization of a quadratic utility function defined over four goods is introduced. In that case, cross price effects are present so that full and partial index number welfare measures are not equivalent. Solutions to this problem are obtained by quadratic programming methods, and cannot easily be reproduced by the reader. However, the message of the results is clear -- the two step valuation method is generally very unreliable.

Zero Cross-Price Effect Examples

Suppose arbitrarily that price-quantity data for a particular good show variation from a maximum price of 25 to a minimum price of 5, and that the underlying demand functions generating the data are q-100/p and q-25-p. The average surplus measures for each case are:[5]

Unit Elastic:

$$\overline{CS} = ((\int_{100}^{p^o} 100/p \; dp) - p^o q^o)/q^o = (100(\ln 100 - \ln p^o) - 100)/(100/p^o).$$

$$= p^o [(\ln 100 - \ln p^o) - 1]$$

where 100 is the price that sets q to one and q^0 is the quantity associated with p^o.[6] For the linear case:

Linear:

$$\overline{CS} = (\int_{25}^{p^o} 25 - p \, dp)/q^o = ((25^2 - 1/2 \; 25^2) - (25p^o - 1/2 \; p^{o^2}))/(25 - p^o)$$

where 25 is the price that sets q to zero and q^0 is the quantity associated with p^0.

For example's sake take three price change situations for each demand specification; two "small" price changes (one far removed from the unit

elastic point of the linear function and one very close to it); and one "large" change over the mid-range of prices (one which validates the $\tilde{\text{LQV}}$ measure).

First, suppose the base price is initially high so p^0=20 and drops by 0.5 to p^1=19.5 after the policy. The absolute value of the arc elasticity of the linear function is 3.76 in this region. For this "small" price change it is obvious that for both functions $\tilde{\text{LQV}}$, $\tilde{\text{PQV}}$ and $\tilde{\text{HCS}}$ will all be approximately equal since there is little difference between the base and post-policy marginal unit values p^0 and p^1 applied to the quantity change. These measures can be compared to the product $\overline{\text{CS}}\cdot\Delta q=\tilde{\text{NB}}$, the approximation $\tilde{\text{CS}}$, and the true value being sought, CS.[7] The results for this first case are shown in table 3.1.

Table 3.1. Case 1: Small Price Change: High Initial Price

Assumed: p^0 = 10; p^1 = 19.5

Calculated:	Linear demand	Unitary elastic
q^0	5.00	5.00
q^1	5.50	5.13
True CS	2.62	2.53
$\overline{\text{CS}}$	2.50	12.18
$\tilde{\text{LQV}}$	10.30(3.82)[a]	2.56(1.01)[a]
$\tilde{\text{PQV}}$	9.75(3.72)	2.50(0.99)
$\text{H}\tilde{\text{CS}}$	9.88(3.77)	2.53(1.00)
$\tilde{\text{CS}}$	2.62(1.00)	2.53(1.00)
$\overline{\text{CS}}\cdot\Delta q$	1.25(0.48)	1.56(0.62)

a. Figures in parentheses are ratios of approximations to true surpluses.

One important lesson of this example is that the $\tilde{\text{LQV}}$, $\tilde{\text{PQV}}$, $\text{H}\tilde{\text{CS}}$ and $\tilde{\text{CS}}$ approximations all work quite well for the unit elastic case when the prize

change is small, as expected, but fail rather dismally for the linear
function because the point of evaluation p^0, q^0 is so far removed from the
point of unit elasticity of this function (p=12.5, q=12.5). Thus, "small"
price changes, in and of themselves, do not guarantee approximation
accuracy using marginal unit values except in the case of unit elasticites.
The second lesson evident from the example is the especially poor
performance of the product of an average surplus and the quantity change,
as expected from Eq.'s 20 and 21.

Next, the same calculations can be made in the neighborhood of the
unit elastic point of the linear function. (Table 3.2) Here, all
approximations except the average unit value method work very well, again
as expected. Even the average value method is fairly good for the unit
elastic demand function, though it still fails for the linear case.

Table 3.2. Case 2. Small Price Change in
Neighborhood of Unit Elastic Point

Assumed: $p^0 = 12.75$; $p^1 = 12.25$

Calculated:	Linear demand	Unitary elastic demand
q_0	12.25	7.84
q_1	12.75	8.16
True CS	6.25	4.00
$\overline{CS}$	6.12	13.51
$L\widetilde{Q}V$	6.37(1.02)[a]	4.08(1.02)
$P\widetilde{Q}V$	6.13(0.98)	3.92(0.98)
$H\widetilde{C}S$	6.25(1.00)	4.00(1.00)
$\widetilde{CS}$	6.25(1.00)	4.00(1.00)
$\overline{CS} \cdot \Delta q$	3.06(0.49)	4.32(1.00)

a. Figures in parentheses are ratios of approximations to true surplus.

Now, for large price changes, suppose $p^0=15.$ If $\tilde{LQV}$ is to be equivalent to $\tilde{CS},$ in the linear case, this implies a $p^1 = p^{1*}$ of 5. The results are given in table 3.3.

Table 3.3. Case 3. Large Price Change for which $\tilde{LQV}=CS$ by Construction for Linear Demand Relation

Assumed: $p^0 = 15;$ $p^1 = 5$

Calculated	Linear demand	Unitary elastic demand
q_0	10.00	6.67
q_1	20.00	20.00
True CS	150.00	109.86
$\overline{CS}$	5.00	13.45
$\tilde{LQV}$	150.00(1.00)[a]	199.95(1.82)
$\tilde{PQV}$	50.00(0.33)	66.65(0.61)
$\tilde{HCS}$	100.00(0.67)	133.20(1.21)
$\tilde{CS}$	150.00(1.00)	133.20(1.21)
$\overline{CS}\cdot\Delta q$	50.00(0.33)	179.28(1.63)

a. Figures in parentheses are ratios of approximations to true surpluses.

Because the price change is non-marginal none of the approximations perform well if the true demand curve is unit elastic. Particularly, the linearization $\tilde{CS}$ overstates the true surplus CS. For the linear case a good result in terms of LQV = $\tilde{CS}$ = CS has been guaranteed by construction, and not much can be said here except that $\tilde{HCS}$ diverges from CS by a factor of 0.67, which is the arc elasticity of the linear schedule. (It is interesting to note that if the same large price change of $10 is initiated at $p^0=1$ instead of 15, this minor alteration of the Case 3 initial conditions breaks the equality of $\tilde{LQV}$ and CS in the linear case.)

All of the above examples can be worked out on a hand calculator, but we next move on to a more complex (quadratic) formulation of the utility function. While the numerical solution of the consumer's utility maximization problem in different price situations requires an optimization algorithm the lessons regarding the questionable usefulness of the various index number approximations remain the same.

A Quadratic Utility Function Example

A useful specification of the consumer's utility function which provides for zero consumption of sane goods in the choice set independent of whether the Lancaster formulation of the household model holds is the quadratic (Pollak, 1971, Wegge 1968). It has received serious consideration in an applied context by Wales and Woodland (1983) and, reflecting on its didactic value, Wegge 1968 observed "... because of the fact that inferior and superior commodities, substitutes and complements, and zero consumption can be allowed for, a quadratic utility indicator seems to be one of the simplest examples which can be used to demonstrate numerically the flexibility of market behavior permissible under the assumption of rational behavior" (p. 222).

Adopting the additive quadratic form for the utility function[8] and a standard neoclassical structure to the problem, assume three recreational activities and a Hicksian composite commodity are in the consumer's choice set. Income is exogenously determined, and units of consumption of all three leisure activities are measured in days with a total leisure time constraint of 125 days. The consumer's optimal choice set can be determined by solving the nonlinear programming problem:

Max U = u(q)

S.t. y ≥ pq

q ≥ 0

$\overset{*}{q}$ ℓ ≤ T

where p is a 1 by 4 row vector of market goods prices, q is a 4 by 1 column vector of quantities with the Hicksian good in the last position and y is a scalar representing income. For the time constraint, T equals 125 and $\overset{*}{q}$ is a 3 by 1 column vector of recreation good quantities (days) and ℓ is a 3 by 1 column vector of 1s.

Figure 3.2 displays typical Marshallian demand schedules for the first activity (call it fishing), at different levels of income, fixing p_2, p_3 and p_4 at 1, 10, and 1 respectively. (To solve the consumer's problem we used Lempke's complementary pivot algorithm (Ravindran 1972).) It is interesting to note that although we might expect nonlinear schedules (Pollak 1971)[9] our demand curves are very nearly linear, suggesting that valuation of a quantity change by an average surplus dollar value will lead to underestimation of the welfare change if a single price changes.

This is indeed the case. In table 3.4 we show the optimal solutions to the programming problem across different income levels for two policy scenarios. The first scenario (I) operates in the high-price low quantity region of the good 1 demand curve, lowering p_1 from a pre-policy level of $13 per day to a post-policy level of $8 with all other prices fixed. The second scenario (II) starts at a pre-policy price of $7, for good 1, reducing it to $2.

While table 3.4 is not particularly interesting in itself, it does demonstrate the zero consumption phenomenon (for good 1 and good 3). Good 2 is an inferior good over all incomes for both price sets, and good 1 is

Figure 3.2. Marshallian Demand for Fishing Days; quadratic utility
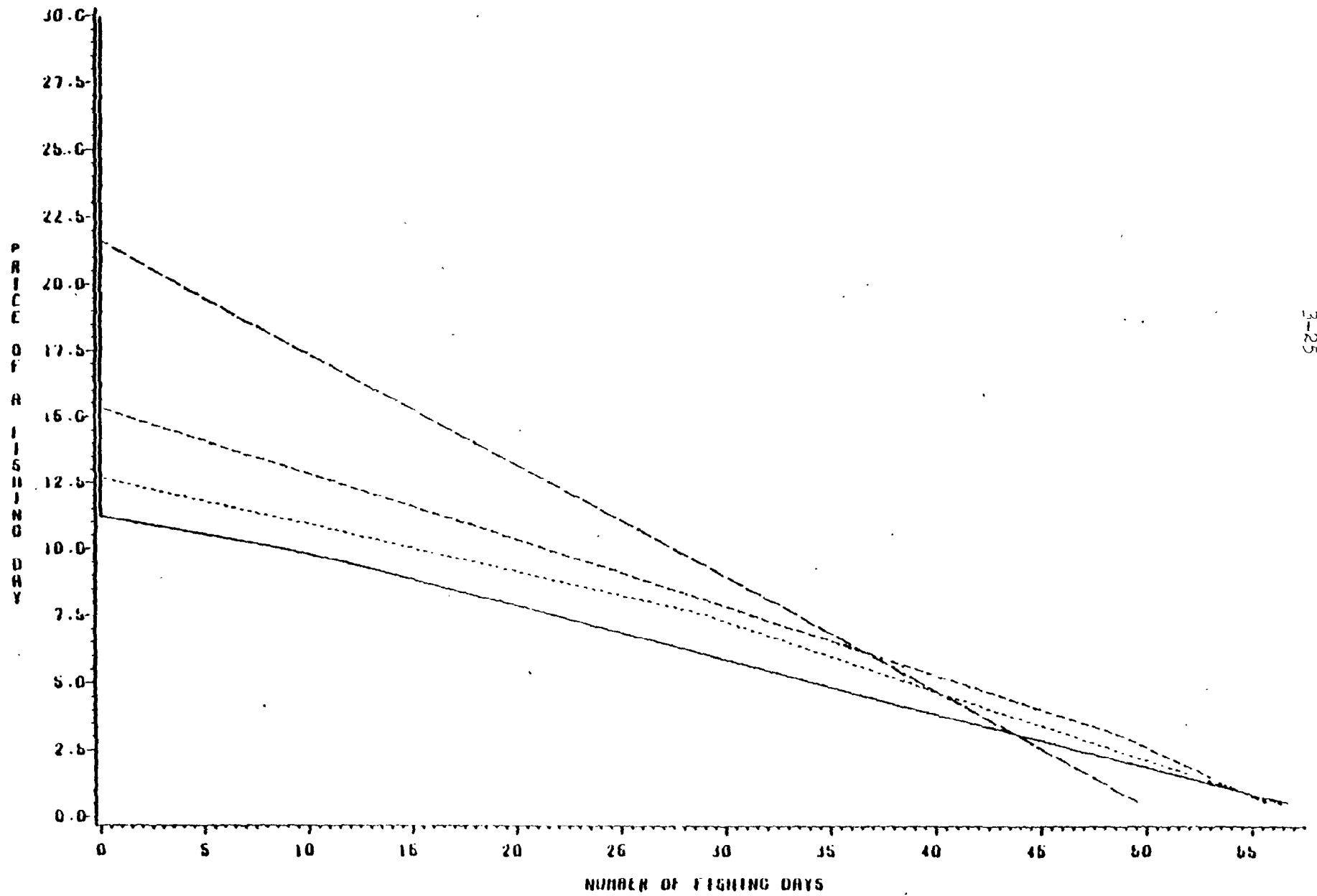
Table 3.4. Optimal Consumption Bundles

Income	Policy Scenario: (p_1)	Fishing Days (q_1)	Passive Days (q_2)	Camping Days (q_3)	Composite Commodity (q_4)	Arc Elasticity o f Demand(q_1)	Quantity Change, q_1	Average Surplus, Good 1(CS)
5000	I. PRE	13	0	119.3	5.7	4823.4			
	POST	8	19.2	105.9	0	4740.9	4.20	19.15	0
	II. PRE	7	24.2	100.8	0	4729.9			
	POST	2	49.4	75.6	0	4825.6	0.61	25.21	2.29
10,000	I. PRE	13	0	113.8	11.2	9774.3			
	POST	a	23.1	101.9	0	9713.2	4.20	23.12	0
	II. PRE	7	27.6	97.4	0	9709.5			
	POST	2	50.0	75.0	0	9825.0	0.52	22.38	2.62
15,000	I. PRE	13	0	108.4	16.6	14725.2			
	POST	8	26.2	97.2	1.6	14677.0	4.20	26.22	0
	II. PRE	1	31.0	94.0	0	14689.1			
	POST	2	50.5	74.5	0	14824.5	0.43	19.53	2.91
20,000	I. PRE	13	3.7	101.3	19.9	19650.7			
	POST	8	27.7	91.1	6.2	19625.2	3.20	23.98	0.39
	II. PRE	7	32.5	89.0	3.5	19648.6			
	POST	2	51.1	73.9	0	19823.9	0.40	18.57	3.39
25,000	I. PRE	13	9.2	93.6	22.2	24564.3			
	POST	8	29.2	85.0	10.8	24573.3	2.18	19.97	1.17
	II. PRE	7	33.2	83.2	8.5	24598.9			
	POST	2	51.7	73.3	0	24823.3	0.39	18.44	4.17

Table 3.4 (continued)

Income	Policy Scenario: (P_1)		Fishing Days (q_1)	Passive Days (q_2)	Camping Days (q_3)	Composite Commodity (q_4)	Arc Elasticity of Demand(q_1)	Quantity Change, q_1	Average Surplus, Good 1(CS)
30,000	I. PRE	13	14.8	85.8	22.5	29477.0	1.48	15.97	2.34
	POST	8	30.7	78.9	15.4	29521.5			
	II. PRE	7	33.9	77.5	13.6	24549.2	0.34	15.74	5.34
	POST	2	49.6	70.5	4.8	29781.7			
35,000	I. PRE	13	20.3	78.0	26.7	34391.3	0.96	11.96	4,29
	POST	8	32.2	72.8	20.0	34469.7			
	II. PRE	7	34.6	71.8	18.6	34499.5	0.26	11.70	7.31
	POST	2	46.3	66.5	12.2	34710.9			
40,000	I. PRE	13	25.8	70.2	29.0	39304.9	0.56	7.95	8.27
	POST	8	33.7	66.7	24.6	39417.8			
	II. PRE	7	35.3	66.0	23.1	39449.9	0.18	7.67	11.32
	POST	2	43.0	62.5	19.6	39656.1			

inferior for income levels beyond $15,000 under price set II. The arc
elasticities of demand for good 1 show that scenario I is confined
principally to the elastic region of each income-specific demand curve, and
scenario II to the inelastic region. The final two columns of table 3.4
contain the information necessary to compute a benefit measure using an
average surplus unit value.[10]

In table 3.5 compensating and equivalent variation (CV, EV) along with
all of the individual-specific benefit measures discussed previously are
displayed for representative consumers,who are distinguished by income
levels and price sets.[11] Three important features of this table illustrate
our previous theoretical results:

(1) Good 1 takes a small share of total expenditure, has a low income
    elasticity, and a true demand curve that is nearly linear. So,
    the Marshallian surplus approximation $\tilde{CS}$ is closely bracketed by
    CV and EV, as expected from Willig, 1976.

(2) The relationship between the partial Harberger measure $\tilde{HCS}$ and the
    Marshallian measure $\tilde{CS}$ is indeed proportional to the arc price
    elasticity, as expected from Eq. (13) above. For example, in the
    first row of table 3.5, $\tilde{HCS}$ is $201.08. With an arc elasticity of
    4.2 from table 3.4, the Marshallian measure from Eq. (13) is
    201.08/4.2=47.88, in this case exactly equal to the value reported
    in column 1, computed independently using Eq. (12).

(3) The measure $\tilde{NB}_i$ obtained by applying an average surplus to the
    quantity change understates the true welfare measure CV by more
    than half, as expected from Eq. 16.

Table 3.6 shows what happens in the aggregate if the example
population of 16 consumers (8 income levels by 2 price scenarios) is

Table 3.5.   Individual-Specific   Monetary   Welfare   Change   Measures

Income and Policy Scenario	Approximate Marshallian and full Harberger Surplus (CS, HCS)	Partial Laspeyres (LQV)	Partial Paasche (PQV)	Partial Harberger (HCS)	Average Surplus (NB)	Compensating Variation (CV)[b]	Equivalent Variation (EV)[b]
500							
I	47.88	248.35	153.20	201.08	0	33.46	33.72
I I	383.93	176.47	50.42	113.44	57.73	183.84	184.22
10,000							
I	57.80	300.56	185.04	242.76	0	46.81	47.14
I I	193.85	156.66	44.76	100.71	58.64	193.75	194.11
15,000							
I	65.55	340.86	209.76	275.31	0	60.99	61.68
I I	203.78	136.71	39.06	87.86	56.83	203.65	204.10
20,000							
I	70.65	311.74	191.84	251.79	9.35	70.20	79.05
I I	204.02	129.90	37.14	83.56	62.95	212.34	212.57
25,000							
I	96.18	259.61	159.76	209.68	23.36	95.71	96.63
I I	212.20	115.08	32.88	73.90	613.55	214.86	215.03
30,000							
I	113.68	207.61	127.76	167.68	37.37	113.29	114.16
I I	208.90	110.18	31.48	70.83	84.05	209.34	208.92
35,000							
I	131.20	155.48	95.68	125.58	51.31	131.00	131.60
I I	202.25	81.90	23.40	52.65	85.53	202.39	202.59
40,000							
I	148.72	103.35	63.60	83.48	65.75	148.54	149.20
I I	195.62	53.69	15.34	34.52	86.82	196.02	195.71

Notes:
a.   All measures theoretically have negative signs for a Welfare improvement, but are reported as absolute values here.
b.   From Golden Section Search, with an interval of numerical uncertainty of $0.01.   In general |CV| < |EV|, except where good 1 is inferior.

(Ratios to CV in parentheses)

Welfare Measure (S)	Low Income Group[a]		High Income Group[a]		Row Total
	Low Price Scenario II	High Price Scenario I	Low Price Scenario II	High Price Scenario I	
Compensating Variation, CV	794 (1.00)	217 (1.00)	830 (1.00)	488 (1.00)	2329 (1.00)
Marshallian and Full Harberger[b] Surplus ($C\tilde{S}$, HCS)	791 (0.996)	250 (1.152)	819 (0.987)	490 1.004	2350 (1.009)
Partial Harberger Surplus ($H\tilde{C}S$)	386 (0.486)	971 (4.475)	232 (0.280)	586 (1.201)	2175 (0.934)
Partial Laspeyres ($L\tilde{Q}V$)[c]	600 (0.756)	1202 (5.539)	361 (0.435)	726 (1.488)	2889 (1.240)
Partial Paasche ($P\tilde{Q}V$)	171 (0.215)	740 (3.410)	103 (0.124)	447 (0.916)	1461 (0.627)
Average Surplus ($\tilde{NB}$)[d]	86 (0.108)	36 (0.166)	363 (0.437)	224 (0.459)	933 to.4011

Notes:
a.   Incomes below median of 22,500 in low income group, incomes above median in high Income group.
b.   See Eq. (10) in text.
c.   Calculated using only Δq, ignoring other quantity changes.
d.   See Eq. (14) in text. **Group-specific** average surpluses, obtained as the mean of the individual average surpluses in each group, applied to the group-specific total quantity change.

partitioned into 4 roughly homogenous, equal sample groups, and group-total welfare measures computed separately for each group and summed. The groups are cross classified by income (less than or greater than 22,500) and price scenario (I, II). So doing arranges individuals in cells in ascending order according to group average arc price elasticity of demand:

Income

Price	Low	High
Low	0.49	0.29
High	3.95	1.30

The patterns of under or overstatement in each column of table 3.6 are consistent with what we would expect from the theoretical development, with one exception. That is, while the $\widetilde{NB}$ measure always understates individually and in the aggregate, the $\overline{HCS}$ measure may come quite close to a proper welfare total.[12] But this can only happen by fortuitous accident, with individuals neatly arrayed across initial and post policy price levels and income levels such that overstatements counterbalance understatements over all.

Summing up, the two step valuation route is dictated by the lack of accurate data on individual marginal willingness to pay for the spectrum of recreation activities. If surveys of population recreation participation contained individual-specific marginal willingness to pay information for potential (as opposed to actually undertaken) visits to all available sites for all leisure related purposes, the two step approach would be

unnecessary. Instead, the welfare change could be obtained directly as the change in the area behind the estimated compensated (or Marshallian) unconditional demand function for visits of a particular sort (Bockstael and McConnell 1983, Morey 1983) just as we would do with a marketed good. But when such price data are not available on an individual-specific level, prices cannot be used in estimation. Instead group average unit values, which are perhaps prices but most likely are not, have to be found to arbitrarily value a quantity change, however estimated.

MODEL SPECIFICATION: THE ROLE OF WATER RESOURCE AVAILABILITY AND POLLUTION VARIABLES IN RECREATION PARTICIPATION EQUATIONS

When unit-day values are used to monetize quantity (days) changes predicted from an econometric recreation participation model under a scenario of recreation resource enhancement, the resultant benefit measure is likely to be inaccurate, no matter how accurate the prediction of quantity change. But prediction accuracy is another fundamental problem with the macro participation equation approach antecedent to, and perhaps as important as, the issue of valuation.

In the case of water pollution control, water resource enhancement is presumed to bring about an augmentation in the quantity of water "suitable" for the activity, decrease the expected congestion at and cost of travel to "suitable" water, and to thereby stimulate an increase in days of participation.   While this chain of reasoning seems plausible it is not universally accepted.   The counter argument is that since such a small fraction of currently available water can be labelled "unsuitable", marginal improvements may, except for localized situations, have an imperceptible impact on participation costs or congestion, and hence on aggregate participation intensity in some pollution-insensitive pursuits

(eg. boating).   A more sophisticated version of this counter-argument is that macro cross-sectional participation models are an inappropriate instrument for identifying pollution impacts.   While there may be sane state level cross sectional variation in the fraction of surface water acreage which is is "unsuitable" for participation, such a measure, being measured too broadly, may not be relevant to individual decision makers. Rather, the physical and chemical characteristics of particular water bodies may well be the appropriate quality attributes influencing choice. And, even if a broad state-level aggregate measure of pollution is relevant to individuals, observed pollution levels in any national cross-section may be everywhere below the threshold levels which affects the perceived utility of a particular kind of recreation experience.   In this case, no demonstrable effect will be revealed in an applied econometric analysis.

While the (negative) effect of pollution levels on the probability and intensity of participation in all kinds of water-based recreation is hardly a universally accepted doctrine, it has been an implicitly maintained hypothesis in many empirical investigations, dating back to the 1966 study by Davidson, Adams and Seneca.   Again, just as in the valuation case, the reason perhaps can be attributed to a paucity of data, rather than a deliberate attempt to guarantee positive benefits of water quality improvements.

To demonstrate how this situation arises, let us assume a linear specification of the participation response function $y = f(\cdot)$, and represent freshwater availability as Q, distance to the nearest marine or Great Lakes coastline as D, the fraction of freshwater area polluted and therefore unsuitable as P_Q; the fraction of marine water polluted as P_M, and let all other influences on the response (y) be collapsed for simplicity into an

augmented intercept, K. Then the participation response function in the general model, ignoring the error specification, is, for individual i:

$$y_i = K_i + \beta_1(Q_i) + \beta_2(Q_i \cdot P_{Qi}) + \beta_3(D_i) + \beta_4(D_i \cdot P_{Mi}) \qquad (22)$$

where $\beta_1 > 0, \beta_2, \beta_3, \beta_4 < 0$. In this representation marine pollution can increase the expected distance of travel to any marine or Great Lakes recreation destination or, otherwise said, the travel-associated cost of participation there. Similarly, freshwater pollution may withdraw freshwater acreage from the perceived pool of available acreage and reduce participation. The extent to which these two pollution effects are perceived and acted upon by the recreationist depends on the magnitude and significance of the parameter estimates of β_2 and β_4 in the econometric model. But that the model in (22) represents a neutral view of the role of pollution, regarding it as a hypothesis that can be tested statistically.

If the investigator happens to be sympathetic to the skeptical view that pollution may have no perceptible influence on participation, the null hypotheses of his restricted model would be $\beta_2 = \beta_4 = 0$. If these restrictions cannot be rejected, the implication is that water pollution reductions are unlikely to produce any direct benefits, although option and existence value benefits, which are not captured by the participation model, cannot be ruled out.

Interestingly enough, tests of this sort are uncommon. Rather, the standard procedure is to posit, as a maintained (i.e., untested, hypothesis, that $\beta_1 = -\beta_2$, and $\beta_3 = \beta_4$. Then, by construction, marine distance is augmented and freshwater acreage reduced by the appropriate pollution fractions <u>prior</u> to estimation, yielding a decidedly "environmentalist" model specification:

$$y_i = K_i + \beta_2 (Q_i (1-P_{Qi})) + \beta_3 (D_i (1+P_{Mi})) \tag{23}$$

where $\beta_1 > 0$, $\beta_2 < 0$. Here, if the parameter estimates of either β_1, β_2 or both are statistically significant, the conclusion that positive benefits will be forthcoming from pollution reductions is inevitable, but perhaps unjustified. [13]

What is one man's reason is another's folly, and unfortunately, statistical criteria cannot always distinguish the two. Because both the skeptical and environmentalist models are restricted versions of the full model in (22) they can be tested separately against it. But, they cannot easily be tested against each other because they are non-nested. So restrictions of the null hypotheses of both of the restricted models may not be rejected in separate tests against the full model. The conundrum raised by the possibility of two plausible but non-nested narrow models is in general irreconcilable, and even if sophisticated non-nested hypotheses testing procedures are undertaken, they may not produce a clear cut decision. While the narrow model with the highest likelihood function value can be taken to represent the preferred specification (Amemiya 1981), this model discrimination criterion (variously labelled the Sargan test or Akaike's Information Criterion) is not really a statistical test with Known properties. Rather, it should be successful "on average" presuming one of the models in the comparison set is indeed the true model.

Realistically, the quality of pollution information obtainable from surveys (as in appendix A to chapter   below) that ask environmental officials questions like "In your state, what fraction of total freshwater surface acreage is unsuitable for activity X due to pollution?" may be too poor to support hypothesis tests of parameter restrictions. [14] If the state cross sectional series on the percentage of water polluted borders on

random noise, or if the overall average has some meaning but across-state differences do not, there is simply not enough information in the pollution data to lead to rejection of the restrictions of either the skeptic or the environmentalist models vis-a-vis the full model. In such a situation, which in our view is commonplace, meaningful tests of the role of pollution in recreation participation models are not possible. Thus it is not surprising that the restrictions of the environmentalist model are maintained hypotheses in many water pollution control benefit studies, especially when those who commission the study, those who undertake it, or both, presume such benefits exist.

SUMMARY AND CONCLUSION

There are several sources of possible error in using the conventional two-step "macro" participation method for approximating a welfare change due to recreational resource enhancement:

(1) Mis-prediction of the change in the quantity demanded due to the policy, as a result of using availability proxies for either price or site attributes (previous chapter).

(2) Error in valuation due to use of either a marginal unit value or an average surplus.

(3) Acceptance of a benefit-producing relationship without testing against a more skeptical null hypothesis.

These problems do not inspire confidence in the two-step method of welfare analysis employed using conventional participation equation models.

Beyond these issues is that of which estimator is most appropriate. This is the subject of the next chapter, which explains why several alternative limited dependent variable estimators are logical candidates

for the estimation of recreation participation models, and why it is often difficult to prefer any one of them over the others.

After this review the estimators are applied to recreation survey data to produce recreation participation models incorporating pollution effects for fishing, and swimming. In those participation equation applications, the restrictions of the environmentalist model are maintained hypotheses, and the benefits of pollution control so produced must be considered with that caveat in mind.

Two separate chapters are then devoted to a slightly different approach to water quality benefit estimation which does not explicitly use the participation equation construct, and there we do test the environmentalist hypothesis. In those chapters we attempt to capture the potential benefits of water quality improvement accruing to the boating category of recreation via the estimated demand for the durable good (the boat) in a first step along with the estimated demand for the activity service flow (boating days) in a second step. The ambiguous results of the environmentalist versus full model hypothesis tests in those chapters suggest the futility of pursuing a similar exercise in participation equation estimation.

NOTES

1.   Some studies (USDI 1973) have used "trip costs" constructed from population survey information in participation equation estimation.   If an unconditional demand function specification is intended, trip costs must be collected on <u>all</u> sites and <u>all</u> possible recreation activities every consumer can choose among.   It is doubtful that trip costs variables constructed by averaging over several trips to many sites in one particular activity category are adequate,   and the problem of missing substitute activity costs because participation in such substitute pursuits is zero is usually impossible to overcome.   An exception is the work of Morey 1981, 1983 who estimates conditional demand functions.

2.   A cursory reading of Deyak and Smith 1978, in both the theoretical and applied sections, leaves the impression that direct travel expenses play no role in reduced form participation models desired from household production theory.   However, such an interpretation is apparently incorrect, since Deyak and Smith specify the marginal cost (shadow price) of service flows as a function of the prices of "recreational market goods" which presumably should include travel cost as a measure of "site price", although they do not so state.

        Notably, the empirical analysis in Deyak and Smith includes no such measure or proxy for it, focusing instead on congestion-type variables measured as the acres of recreational facilities per capita.   Thus, their econometric model specification appears to be distinct from their theoretical model.   Our previous work, which followed Deyak and Smith's empirical (not theoretical) specification appears deficient in this regard

(Vaughan and Russell 1982) as are several other empirical analyses of recreation participation in the literature.

The omission has rarely been explicitly addressed until recently, when Mendelsohn and Brown 1983 observed "In order to assess the usefulness of the household production function it is important to remember that the fundamental purpose of recreation analysis is to determine the value of the quality and quantity of the public good, the recreation site. The recreation site is a good which enters like other goods as an input into the household production function. The critical issue is to value the site or its objective qualities in terms of the price of the site or the price of each quality.... Although the household production function may be able to provide insights about why people exhibit certain tastes for goods (sites) the tool is an unnecessarily cumbersome approach to measure the value of sites or their qualities" (pp. 610-611).

3.   When the constant elasticity demand curve exhibits unitary elasticity formula (17) is indeterminate. But the limit of $\widetilde{NB}/NB$ as n approaches one can be calculated by L'Hôpital's rule as (k-1) **(1-lnq)/(-lnk).**

4.   The constant unit elastic demand specification is theoretically consistent with a Cobb-Douglas utility function. In this case the share of total income allocated to the good in question is a constant as is the dollar amount spent on it, irrespective of price, since the product of price and quantity is a constant. The linear specification can be regarded as an arbitrary first order approximation to the constant unit-elastic function.

5.   The unit elastic formula is derived in Varian 1978, p. 213.

6.   For a derivation of this formula see Vaughan and Russell 1982.

7. All formulas used for these computations are reported elsewhere in the text, except for CS which in the unit elastic case is $\int_{p_0}^{p_1} 100/p \, dp = 100$ $(\ln p_1 - \ln p_0)$. In the linear case $\tilde{CS}$ is equal to CS.

8. The additive quadratic utility function is not both globally quasiconcave and nondecreasing, so a satiation point (bliss) can be reached, marginal utility can be negative, and the own Slutsky substitution effects can become positive (compensated demand curves can become upward sloping beyond bliss). Yet the additive quadratic utility function is quasisconcave and nondecreasing over a subset of the commodity space--the region southwest of bliss--which is the region of the "economic" problem of choice. The additive form of the general quadratic utility function, useful for didactic purposes, is defined (Pollak 1971)

$$U(q) = - \sum_i c_i (d_i - q_i)^2$$

where c_i and d_i are positive parameters. The cardinal properties of the additive quadratic (Philips 1974) are linear marginal utilities $(\partial U/\partial q_i = 2c_i d_i - 2c_i q_i)$ which can become negative for sufficiently large q_i; diminishing marginal utility $(= -2a_i)$ and independence $(\partial^2 U/\partial q_i \partial q_j = 0)$.

The parameter values used in the example are $c_i = 1$ for all i and $d_i = 182$, $d_2 = 183$, $d_3 = 204$ and $d_4 = 49{,}000$.

9. The demand curves derived by Pollak 1971 are inherently nonlinear and convex to the origin while the demand curves from the programming solution in Wegge 1968 and our results are piecewise linear and concave to the origin. The discrepancy arises because Pollak 1971 ignored the non-negativity quantity constraint and derived the demand curves using standard Lagrangian techniques while the programming solution reflects the operation of the inequality and non-negativity constraints.

10. The individual-specific average value column mimics the response elicted in the 1975 National Survey of Hunting, Fishing and Wildlife-Associated Recreation, which asked "Having thought about how much this activity cost you in 1975, <u>how much more</u> money would you spend annually on your favorite activity before deciding to stop doing it because it is too expensive?" In our model, this value is captured as the average of the equivalent and compensating variations between the base level of participation in fishing and the zero level divided by the base level days of fishing.

11. Policy benefits can be calculated both in compensating variation (CV) and equivalent variation (EV) terms. (Deaton and Muellbauer 1980). Define

$$CV = e(p^1, u^0) - e(p^0, u^0)$$

$$EV = e(p^1, u^1) - e(p^0, u^1)$$

where $e(\cdot)$ represents the minimal expenditure required to reach the stated utility level, given the price vector. Obviously, $e(p^0, u^0) = e(p^1, u^1) = y^0$ if the consumer's income constraint is binding (he is not beyond bliss).

To obtain the correct EV and CV measures, the quadratic programming model must be resolved (parameterized) in steps away from either the pre or post policy solution, where the parameterization involves decrements (for CV) or increments (for EV) in the income available below or above y^0. This is necessary because the expenditure function cannot be derived analytically. Instead, a Golden Section search algorithm (Biles and Swain 1980) was employed to find CV and EV numerically. Specifically, the problem for CV is to find (by numerical search) the income level $y_C^* < y^0$ that, under price vector p^1, allows the consumer to obtain the optimal pre-policy utility level u^0 with income y^0. Of course, u^0 is known from the pre-policy optimization run with price vector p^0. Then $CV = y_C^* - y^0$.

The logic for the EV calculation is similar. All welfare measures are reported as absolute values.

12. Table 3.5 uses an individual-specific average surplus, $\overline{CS}^i$. Whether or not <u>total</u> benefits over all individuals differ much if individual changes are valued with an individual value $\overline{CS}^i$ and summed or the total quanity change is valued using an average of the average unit values $\overline{CS}$ depends on the correlation between changes in quantities and $\overline{CS}^i$. If the correlation is positive use of $\overline{CS}^i$ instead of $\overline{CS}$ produces a lower total welfare change, and vice versa if the correlation is negative.

13. See Vaughan and Russell 1982 for an example of this sort of specification, which was invoked without scrutiny following Davidson, Adams and Seneca 1966.

14. For example, state officials were asked this sort of question regarding the percentage of fishable water by Vaughan and Russell 1982. That study also employed a mathematical water quality simulation model along with rules translating the water quality model's ambient water quality measures into fishable water to predict the latter as a fraction of total freshwater. The unweighted coefficient of determination between the survey series and the synthetic series was only 0.31, and when data were weighted by acreage, it dropped to a disappointing 0.08.

REFERENCES

Amemiya, Takeshi. 1981. "Qualitative Response Models: A Survey," <u>Journal of Economic Literature</u> vol. 19 (December) pp. 1483-1536.

Becker, G. S. 1965. "A Theory of the Allocation of Time," <u>Economic Journal</u> vol. 75, no. 299 (September) pp. 493-517.

Biles, William E. and James J. Swain. 1980: <u>Optimization and Industrial Experimentation,</u> (New York: John Wiley and Sons).

Bockstael, Nancy E. and Kenneth E. McConnell. 1983. "Welfare Measurement in the Household Production Framework," <u>American Economic Review</u>, vol. 74, no. 4 (September), pp. 806-814.

Cicchetti, Charles J., Anthony C. Fisher, and V. Kerry Smith. 1973. "Economic Models and Planning Outdoor Recreation," <u>Operations Research</u> vol. 21, no. 5 (September-October) pp. 1104-1113.

_____. 1973. <u>Forecasting Recreation in the United States</u> (Lexington, Mass.: Lexington Books).

_____ and V. Kerry Smith. 1973. "Interdependent Consumer Decisions: A Production Function Approach," <u>Australian Economic Papers</u>, vol. 11 (December), pp. 239-252.

_____ and _____. 1976. <u>The Costs of Congestion</u>, (Cambridge: Ballinger) Ch. 2.

Davidson, Paul, F., Gerald Adams, and Joseph Seneca. 1966. "The Social Value of Water Recreational Facilities Resulting from an Improvement in Water Quality: The Delaware Estuary," in Allen V. Kneese and Stephen C. Smith; eds., <u>Water Research</u> (Baltimore, Md.: Johns Hopkins University Press for Resources for the Future).

Deaton, Angus and John Muellbauer. 1980. <u>Economics and Consumer Behavior</u> (Cambridge: Cambridge University Press), Ch. 10.

Deyak, Timothy A. and V. Kerry Smith. 1978. "Congestion and Participation in Outdoor Recreation: A Household Production Function Approach," <u>Journal of Environmental Economics and Management</u> vol. 5, no. 1 (March) pp. 63-80.

Dyson, Pamela J. 1984. <u>Recreational Water Availability in the United States: The Impact of Pollution Control</u>. (Washington, D.C.: Resources for the Future).

Dwyer, John F. John R. Kelly and Michael D. Bowes. 1977. "Improved Procedures for the Valuation of the Contribution of Recreation to National Economic Development", (Urbana: University of Illinois Water Resources Center) Research Report No. 128.

Freeman, A. Myrick III. 1979b. The Benefits of Environmental Improvement. (Baltimore, Md.: Johns Hopkins University Press for Resources for the Future).

_____. 1982. Air and Water Pollution Control (New York: John Wiley and Sons), 186 pp.

Greig, Peter J. 1983. "Recreation Evaluation Using a Characteristics Theory of Consumer Behavior," American Journal of Agricultural Economics, vol. 65, no. 1, (February), pp. 90-97.

Hay, Michael J. and Kenneth E. McConnell. 1979. "An Analysis of Participation in Nonconsumptive Wildlife Recreation," Land Economics vol. 55 (November) pp. 460-471.

Ironmonger, D. S. 1972. New Commodities and Consumer Behavior, (Cambridge: Cambridge University Press), Ch. 1-4.

Lancaster, Kelvin J. 1966. "A New Approach to Consumer Theory." Journal of Political Economy, vol. 74, no. 2 (April), pp. 132-157.

Mak, James and James E. Moncur. 1980. "The Choice of Journey Destinations and Lengths of Stay: A Micro Analysis," Review of Regional Studies, vol. 10, no. 2 (Fall) pp. 38-48.

McConnell, Kenneth E. and Ivar Strand. 1981. "Measuring the Cost of Time in Recreation Demand Analysis: An Application to Sportfishing," American Journal of Agricultural Economics, vol. 63, no. 1 (February), pp. 153-156.

McFadden, Daniel and Fred Reid. 1975. "Aggregate Travel Demand Forecasting from Disaggregated Behavioral Models," Transportation Research Record, vol. 534, pp. 24-37.

McKenzie, George W. 1983. Measuring Economic Welfare: New Methods, (New York: Cambridge University Press), 187 pp.

Mendelsohn, Robert and Gardner M. Brown, Jr. 1983. "Revealed Preference Approaches to Valuing Outdoor Recreation," Natural Resources Journal, vol. 23 (July), pp. 607-618.

Morey, Edward R. 1981. "The Demand for Site-Specific Recreational Activities: A Characteristics Approach," Journal of Environmental Economics and Management, vol. 8 (December), pp. 345-371.

_____. 1983. "Characteristics, Consumer Surplus, and New Activities: Theory and Evidence," Discussion Paper, Department of Economics, University of Colorado.

Pearce Bowden Economic Consultants Limited. 1971. "The Value of Fresh Water Sport Fishing in British Columbia," Victoria, British Columbia: Department of Recreation and Conservation.

Phlips, Louis. 1974. Applied Consumption Analysis, (Amsterdam: North Holland).

Pollak, Robert A. 1971. "Additive Utility Functions and Linear Engel Curves," Review of Economic Studies, vol. 38, no. 4, (October), pp. 401-414.

Pollard, J. H. 1971. "On Distance Estimators of Density in Randomly Distributed Forests." Biometrics, vol. 17 (December), pp. 991-1002.

Quandt, Richard E., ed. 1970. Demand for Travel: Theory and Measurement, (Lexington, D.C.: Heath), 304 pp.

Rausser, Gordon C. and Ronald A. Olivera. 1976. "An Econometric Analysis of Wilderness Area Use," Journal of the American Statistical Association, vol. 71, no. 354, pp. 276-285.

Ravindran A. 1972. "Algorithm 431: A Computer Routine for Quadratic and Linear Programming Problems [H], Communications of the Association for Computing Machinery, vol. 15, pp. 818-820.

Rugg, Donald. 1973. "The Choice of Journey Destination: A Theoretical and Empirical Analysis," Review of Economics and Statistics, vol. LV, no. 1, (February), pp. 64-72.

Russell, R. Robert and Maurice Wilkinson. 1979. Microeconomics, (New York: John Wiley and Sons).

Sorg, Cindy, John Loomis, Dennis M. Donnelly and George Petersen. 1984. "The Net Economic Value of Cold and Warm Water Fishing in Idaho," (unpublished working paper).

U.S. Department of the Interior, Bureau of Outdoor Recreation. 1973. Outdoor Recreation: A Legacy for America (Washington, D.C.), Appendix A.

U.S. Department of the Interior, n.d. 1975 National Survey of Hunting, Fishing and Wildlife - Associated Recreation (Washington, D.C.).

U.S. Department of the Interior, Heritage Conservation and Recreation Service. 1979. The Illinois Wild and Scenic River Study (Washington, D.C.).

Varian, Hal. R. 1978. Microeconomic Analysis (New York: W. W. Norton).

Vaughan, William J. and Clifford S. Russell. 1982: Freshwater Recreational Fishing, the National Benefits of Water Pollution Control. (Resources for the Future: Washington, D.C.)

_____ and _____. 1984. "The Role of Recreation Resource Availability Variables in Recreation Participation Analysis," Journal of Environmental Management, vol. 19, pp. 185-196.

Wales, T. J. and A. D. Woodland. 1983. "Estimation of Consumer Demand systems with Binding Non-Negatively Constraints," Journal of Econometrics, vol. 21, no. 3 (April), pp. 263-285.

Water Resource Council. 1979. "Procedures for Evaluation of National Economic Development (NED) Benefits and Costs in Water Resources Planning (Level C), Final Rule," Federal Register, vol. 44, no. 242, December 14, 1979, pp. 72892-977.

Wegge, L. 1968. "The Demand Curves from a Quadratic Utility Indicator," Review of Economic Studies, vol. 35, no. 102 (April), pp. 209-224.

Wennergren, W. Boyd. 1967. "Surrogate Pricing of Outdoor Recreation," Land Economics, vol. 43, no. 1 (February), pp. 112-116.

Willig, Robert D. 1976. "Consumer's Surplus Without Apology," American Economic Review, vol. 66, no. 4 (September) pp. 589-597.

Ziemer, Rod, F., Wesley N. Musser, Fred C. White and R. Carter Hill. 1982. "Sample Selection Bias in Analysis of Consumer Choice: An Application to the Warmwater Fishing Demand," Water Resources Research, vol. 18, no. 2 (April) pp. 215-219.

Chapter 4

ESTIMATION OF QUALITATIVE AND LIMITED
DEPENDENT VARIABLE MODELS

In this chapter we discuss some problems that arise in the econometric estimation of participation models. This material may be considered as complementary to that in chapter 3 on model specification. The treatment will be quite detailed but even so will only brush the surface of a rich and rapidly growing literature.

However, several high quality surveys are available for the reader who wishes to pursue the matter more deeply. The 1981 and 1984 surveys by Amemiya are excellent overviews of qualitative and limited dependent variable models, respectively, and the 1983 monograph by Maddala provides broad coverage in both these areas. The often-cited 1981 volume edited by Manski and McFadden is also an excellent survey of topics in qualitative and limited dependent variable estimation.

Some definitional preliminaries are appropriate here. First, standard practice is followed and random variables represented in upper-case notation, their realizations in lower-case. Second, the terms "censored distribution" and "truncated distribution" will be used with considerable frequency. The introduction to chapter 6 of Maddala (1983) provides a good heuristic explanation of censoring and truncation as they pertain to the normal econometric model. For completeness, we present two more formal explanations of these two phenomena as found in the statistical literature.

First, Kendall and Stuart (1973) describe truncation and censoring as follows, using their now-classic "target" example:

Suppose first that the underlying variate x simply cannot be observed in part or parts of its range. For example, if x is the

distance from the centre of a vertical circular target of fixed radius **R** on a shooting range, we can only observe **x** for shots actually hitting the target. If we have no knowledge of how many shots were fired at the target (say, **n)** we simply have to accept the **m** values of **x** observed on the target as coming from a distribution ranging from 0 to R. We then say that the distribution of x is truncated on the right at R. Similarly, if we define y in this example as the distance of a shot from the vertical line through the centre of the target, y may range from -R to +R and its distribution is doubly truncated. Similarly, we may have a variate truncated on the left (e.g. if observations below a certain value are not recorded). Generally, a variate may be multiply truncated in several parts of its range simultaneously. A truncated variate differs in no essential way from any other but it is treated separately because its distribution is generated by an underlying untruncated variable, which may be of familiar form.

On the other hand... suppose that we know how many shots were fired at the target. We still only observe **m** values of **x,** all between 0 and R inclusive, but we know that n-m = r further values of x exist, and that these will exceed R. In other words, we have observed the first m order-statistics $x_{(1)}, \ldots, x_{(m)}$ in a sample of size n. The sample of x is now said to be censored on the right at R. (Censoring is a property of the sample whereas truncation is a property of the distribution.) Similarly, we may have censoring on the left (e.g. in measuring the response to a certain stimulus, a certain minimum response may be necessary in

order that measurement is possible at all) and double censoring, where the lowest r_1 and the highest r_2 of a sample of size n are not observed, only the $m=n-(r_1+r_2)$ being available for estimation purposes. (Kendall and Stuart (1973), p. 541).

A second explanation is that of Johnson and Kotz (1969), who note that:

There is clearly a close analogy between censoring and truncation, but the differences are evident. Censoring modifies the selection of the random variables; truncation directly modifies the distribution. In other words, censoring is an agreement to ignore observed values because they are larger (or smaller) than a certain number of other observed values, while truncation is omission of values outside predetermined, fixed, limits. (Johnson and Kotz (1969), p. 27).

It should also be noted at the outset that the following discussion of estimation techniques for quantitative dependent variables (e.g. measures like time, days, number of trips, etc.) does not deal with the system or multi-activity structure in terms of which recreation participation models might be cast. That is, one can easily conceive of a system of recreation participation models (fishing, boating, swimming) analogous to more familiar systems of demand equations (food, drink, shelter, and clothing, for example) discussed in the econometrics literature and estimated by techniques such as seemingly unrelated regressions. However, although there exist systems estimation techniques for limited dependent variable models of the nature assessed below (see, e.g., Wales and Woodland (1983)), such techniques are expensive and not easily implemented. Estimation techniques for the single-equation or single-activity models discussed in

this chapter are far more easily and relatively less expensively implemented, and, as such, the discussion to follow is confined to those models that can reasonably be estimated within the scope of this project.

We also elect to set aside for future research consideration of models of the sort discussed by Dubin and McFadden (1984) and Hanemann (1984), these concerned in part with situations wherein individuals select one good or activity from a set of k possible goods or activities. Although such research has potentially fruitful applications in the analysis of recreation participation decisions, full treatment is beyond the scope of this chapter.

The plan for the remainder of this chapter is as follows. First, we briefly assess problems associated with least squares estimation of participation models. Then we turn to a discussion of some techniques that might be considered more or less appropriate for the estimation problems attendant to recreation participation analysis. Following this we turn to a discussion of prediction based on the estimation of the various models. A summary concludes the chapter.

SOME PROBLEMS WITH LEAST-SQUARES ESTIMATION OF PARTICIPATION

The data used in participation analysis commonly displays one or more properties that make simple least squares inappropriate, because the resulting parameter estimates are biased and inconsistent, The alternative techniques usually involve iteration and are more costly than the simple, familiar methods. To see the origins of the problems consider the multivariate linear model:

$$Y_i = X_i \beta + \varepsilon_i, \text{ where } i \text{ indexes observations} \tag{1}$$

and ε_i has zero mean and constant finite variance σ^2. The model satisfies

full ideal conditions (Schmidt p. 2) when:

i) X is a nonstochastic matrix of rank k<T, and has the property that

$$Q = \lim_{T \to \infty} X'X/T \text{ is finite and nonsingular;}$$

ii) $\varepsilon = [\varepsilon_i]$ is distributed multivariate $N(0, \sigma^2 I_T)$.

But whether or not ε is distributed normally, it can be shown that the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ is unbiased and consistent.

As discussed in detail below, a very general characterization of quantitative participation data is that it is data bounded from below by zero, i.e. it is realized only in nonnegative quantities. Of specific concern here are measures like "amount of time spent engaged in some activity." Such measures are generally modeled econometrically as the censored or truncated counterparts of normally-distributed latent random variables Y_i^* having $E(Y_i) = X_i\beta$, $Var(Y_i) = \sigma^2$. However, if the realizations of Y_i are censored from below at zero, we have

$$E(Y_i^* | y_i > 0) = X_i\beta + \sigma\phi_i/\Phi_i, \tag{2}$$

$$E(Y_i^*) = X_i\beta\Phi_i + \sigma\phi_i,$$

where ϕ_i and Φ_i are the standard normal density and distribution functions evaluated at $(X_i\beta/\sigma)$.[1] In the truncated case, where $Pr(y_i 1 > 0) = 1$,

$$E(Y_1^*) = X_i\beta + \sigma\phi_i/\Phi_i. \tag{3}$$

The problems inherent in least squares estimation may be explained using these expectations. $E(\sigma\phi_i/\Phi_i) \neq 0$, then $E(\varepsilon_i) \neq 0$ so that ε_i is defined as the difference between either $E(Y_i)$ or $E(Y_i | y_i > 0)$ and $X_i\beta$ in (2). Thus least squares regression of y on X will yield inconsistent estimates of β, because the null error expectation assumption has been violated. (Heckman (1976) provides a good general discussion of such problems.)

Although not all measures of interest in our analysis are cast in terms of normally-distributed, partially-observed, random variables, these constitute the main realm of our inquiry. In the other cases we shall investigate, however, there are other characteristics of the data or statistical distributions assumed that render least squares inappropriate, given the objective of consistent parameter estimation. For example, least-squares estimation strategy is generally completely inappropriate when outcomes are qualitative, as no objective function of interest can be cast in terms of linear expectations functions like those above. We now turn to an assessment of various approaches to the estimation of participation models.

TOBIT PARTICIPATION MODELS

A logical starting point in any discussion of limited dependent variable model estimation is the basic Tobit model. The nature of several of the participation measures of interest in the micro data sets being analyzed in this study is such that Tobit estimation would seem--at least at first blush--to be a sensible approach.

Tobit estimation has been utilized in a variety of areas in applied microeconomics, ranging from labor supply (see the excellent survey by Killingsworth (1983)), to health economics (Ostro, (1983)), to commodity demands or expenditures (Tobin (1957), Pitt (1983)), and many others (see Amemiya (1984) for an extensive bibliography).

The basic idea underlying Tobit estimation is that one posits the existence of (latent) normally, independently-distributed (NID) random variables $Y_i^* \sim NID(X_i \beta, \sigma^2)$. In many interpretations of the Tobit model, the Y_i^*, are stochastic indicators of intensity of desire for undertaking

some activity. Owing to the nature of the activity, however, some realizations of the Y_i^* are censored while for the others, the intensities, are mapped directly into actual undertakings of the activity. Some threshold, in effect, is crossed such that the activities are actually undertaken. For example, the fundamental idea behind Tobin's seminal paper is that the Y_i^* represent intensities of desire to purchase consumer durables. When certain (assumed known) thresholds are crossed, these intensities become actual purchases: In most applied areas, the thresholds are zero, so that the mappings from intensities into undertaken activities can be looked at as occurring when the realizations of the Y_i^* occur in the interior of commodity space. Otherwise corner solutions obtain (for one discussion of estimation in the Kuhn-Tucker/corner-solution/Tobit context, see Wales and Woodland (1983)).

Assuming, then, that the thresholds are known and constant across individuals, the basic Tobit model can be described by (4):

$$Y_i^* \sim NID(X_i \beta, \sigma^2)$$
$$y_i = \max(C, y_i^*).$$

$$(4)$$

Setting $C = 0$ gives the model we shall discuss below. Letting Ω_0 signify the index set for observations for which $\max(0 \quad y_i^*) = 0$, and Ω_1 be the index set for observations for which $\max(0, y_i^*) > 0$, then the likelihood, function for the Tobit model described here is

$$L = \prod_{i \varepsilon \Omega_0} (1 - \phi(\frac{X_i \beta}{\sigma})) \prod_{i \varepsilon \Omega_1} (\phi(\frac{y_i - X_i \beta}{\sigma})/\sigma),$$

$$(5)$$

where ϕ is the standard normal distribution function and $\phi(Z) = d\phi(Z)/dZ$. In log form (5) is

$$\ell = \sum_{i \epsilon \Omega_0} \ln(1-\Phi_i) - |\Omega_1| \ln\sigma - \sum_{i \epsilon \Omega_1} \ell n \phi_i \qquad (6)$$

where $|\cdot|$ denotes cardinality and where terms not involving (β, σ) are dropped.

The first-order conditions for maximizing ℓ are the $(k + 1)$ equations

$$\partial\ell/\partial\beta = \sum_{i \epsilon \Omega_0} (-\lambda_i/\sigma)X_i' + \sum_{i \epsilon \Omega_1} (y_i - X_i\beta)X_i'/\sigma^2 = 0$$

$$\partial\ell/\partial\sigma = \sum_{i \epsilon \Omega_0} \lambda_i(X_i\beta)/\sigma^2 + \sum_{i \epsilon \Omega_1} ((y_i - X_i\beta)^2 - \sigma^2)/\sigma^3 = 0, \qquad (7)$$

where $\lambda_i = \phi_i/(1-\Phi_i)$. Using terms in these equations, the method of Berndt-Hall-Hall-Hausman (1974) among others, can be used for optimization, and statistical inference is based on the asymptotic t-tests generated by utilizing $[\sum_{i=1}^{N}(\ell_i \ell_i')]^{-1}$ as the estimate of $\text{cov}(\hat{\beta})$ (ℓ_i is the i-th term of $[(\partial\ell/\partial\beta)', [\partial\ell/\partial\sigma)]')$.

Several characteristics of the Tobit model are noteworthy. First, as Amemiya (1984) points out, the likelihood function (5) can be rewritten as

$$L = [\prod_{i \epsilon \Omega_0}(1-\Phi_i) \prod_{i \epsilon \Omega_1} \Phi_i] [\prod_{i \epsilon \Omega_1}(\phi_i/\Phi_i\sigma)] \qquad (8)$$

Written in this form, the likelihood function of the Tobit model can be viewed as the product of the likelihood functions of a binomial probit model with parameter vector $\alpha = \therefore$ (first brackets) and a truncated-at-zero normal distribution with parameters (β, σ) and $E(Y_i) = X_i\beta$ $\sigma\phi_i/\Phi_i$ (second brackets).[2] As such, separate maximization subject to the restrictions that the probit parameter vector be a positive scalar multiple (specifically $1/\sigma$) of the parameter vector of the truncated normal model yields the Tobit model. The probit component be viewed as the model of

whether or not the threshold is crossed, while the truncated normal component models the conditional phenomenon of the magnitude of the activity given that the activity is undertaken.

It is certainly reasonable to consider the possibility that the parameter restrictions described in the preceeding paragraphs are in fact invalid. If they were, it would indicate that the model of threshold crossing is not as intimately related to the conditional model of the magnitude of the undertaken activity as is implied by the Tobit model. In the context of recreation participation, this could mean that the decision about whether or not to engage in some form of water-based activity is governed by a set of parameters different than that determining the amount of participation undertaken given that some participation occurs. We discuss such issues in greater detail later in the chapter.

Another characteristic of the Tobit model that merits discussion is the fact that the parameters estimated under one assumptions of the Tobit model are in general nonrobust to departures from many of the underlying assumptions. That is, violation in the data of some of the properties implied when the likelihood function is written in the form (5) will lead to inconsistent estimates of the parameters (β, σ). This phenomenon, is common in many types of models that are estimated by means of maximum likelihood.

Two of the most often discussed violations that bode dire consequences for Tobit parameter estimates are violations of the NID assumption. First, note that normal, homoscedastic errors are implied when writing the likelihood function in the form (5). Two possible violations of this assumption are that the error variances are nonconstant across observations, and second, that the error structure, though perhaps

homoscedastic, is nonnormal. The results of several studies, summarized by Amemiya (1984), indicate that under either type of departure, the maximum likelihood Tobit parameter estimates are inconsistent.

It is, of course, generally unknown ex ante whether or not the data being analyzed are characterized by the ideal properties. It then becomes essential to determine whether there exist such violations if one is to have some degree of confidence in the consistency properties of the Tobit parameter estimates. We describe briefly two tests that have been proposed to detect departures from the Tobit "ideal" conditions. The first test is for heteroscedasticity of a given form, while the second is a more general test for misspecification.

The idea behind the test for heteroscedasticity, proposed by Smith and Maddala (1983), can be motivated as follows. The (k+1)st first order condition of the Tobit ML model, $\partial \ell / \partial \sigma = 0$, can trivially be rewritten $(\partial \ell / \partial \sigma)(\partial \sigma / \partial \theta) = 0$, if we assume that $\sigma = \sigma(\theta) = \theta_0$, where the dimensionality of θ is one, and that $\partial \sigma / \partial \theta = 1$. In general, however, it is possible that the dimension of θ is greater than one and that σ follows a perhaps complicated parametric relationship that can vary across observations. In what follows, we consider the case where $\sigma_i = z_i \theta = \theta_0 + z_{i1} \theta_1$, where z_i is some proper or improper subset of X_i. Homoscedasticity implies $\theta_1 = 0$.

In this context, the origin of the inconsistency of the Tobit estimates under heteroscedasticity is as follows. In assuming homoscedasticity, the analyst estimates the parameters (β, σ) based on the (k+1) likelihood equations (7). Given heteroscedasticity of the above form, not only do these equations depend on θ_1, as well as (β, θ_0), but the p likelihood equations $(\partial \ell / \partial \sigma)(\partial \sigma / \partial \theta_1) = 0$, where $p = \dim(\theta_1)$, are entirely

omitted from estimation. Given this, the inconsistency is hardly surprising.

Smith and Maddala propose a simple test for heteroscedasticity when the $\sigma_i = Z_i \theta$ hypothesis seems a reasonable alternative to the $\sigma_i = \theta_0$ (homoscedasticity) hypothesis. The test is simply to base estimation on the $(k+p+1)$ likelihood equations

$$(\partial \ell / \partial \beta) = 0$$
$$(\partial \ell / \partial \theta_0) = 0 \qquad\qquad (9)$$
$$(\partial \ell / \partial \theta_1) = 0$$

The test for heteroscedasticity, then, is a likelihood ratio test based on the restriction $\theta_1 = 0$. One can also examine the asymptotic t-statistics on the individual elements of θ_1 to see if any of the hypotheses $\theta_{1m} = 0$, $m = 1,\ldots,p$, can be rejected.

The second and more general test for misspecification of the Tobit model is that proposed by Nelson (1981). Because of its generality, it is both valuable and nondiagnostic. It is valuable because the analyst need not specify the nature of the suspected departure from the Tobit assumptions. It is nonillustrative because, as an omnibus test, should misspecification be indicated the source thereof is not made apparent. The test in principle can detect problems such as errors in measurement on the dependent variable (Stapleton and Young (1984)), nonnormality, and heteroscedasticity, but can also detect other phenomena such as omitted variables. As such, a significant Nelson statistic is important, but still leaves the researcher in somewhat of a quandary.

The Nelson test is a Hausman (1978) test based on moment estimators of functions of the model's parameters which are consistent but inefficient under general conditions. The details are fairly complicated, and the

reader is referred to Nelson's work for their development. In the discussion of truncated models below, an extension of Nelson's test is proposed, and some details of the basic Nelson procedure are discussed in that development.

CRAGG-CLASS PARTICIPATION MODELS

In a 1971 paper, Cragg proposed a set of models for situations that can be depicted as follows. An economic agent makes two decisions. A dichotomous decision is made about whether or not to engage in some activity. Conditional on an affirmative for this decision, a decision is made regarding how much of the activity to pursue. The activities can be construed in the broadest of terms: expenditures, quantities demanded or supplied, or the amount of time spent in recreation participation. Such models have come to be known as "hurdles" models, that is, conditional or some hurdle being crossed, a decision is made about sane magnitude of interest. Although these decision processes might in some cases seem logically to be ordered in a temporal manner, the statistical properties of the model abstract from any temporal considerations, the quantity decision being described in terms of conditional densities.

Cragg proposed several models. However, because of the nature of the present study, only two members of this set will concern us here, these being the formulations wherein the quantity or second-stage decision is defined only for positive real. This is in obvious reference to ideas like "given that an individual participated in activity x, how much time was spent engaging in the activity." Although Cragg's other formulations are also interesting, their discussion is omitted for economy of space.

For notational ease, we will assume that the same vector of independent variables, X_i, influences both the first-and second-stage decisions. This is a completely innocuous assumption, however, as elements of parameter vectors can be restricted equal to zero to accommodate more general cases. Regardless of the specification of the second-stage or conditional decision, the first-stage is described by a binary probit model, i.e. the existence of latent random variables $Y_{i1}^* \sim N(X_i \beta_1, \sigma_1^2)$ is posited. Only the signs of the realizations are recorded, however, and are codified according to

$$y_{i1} = 1, \quad y_{i1}^* \geq 0 \qquad\qquad (10)$$

$$\phantom{y_{i1}} = 0, \quad y_{i1}^* < 0$$

Because of this codification scheme, there is no information about the scale of the random variables Y_{i1}^* (i.e. the mappings of y_{i1}^* into y_{i1} are unaffected by transformations of Y_{i1}^* the form θY_{i1}^* for $\theta > 0$). Therefore, some normalization is required, the most common being $\sigma_1 = 1$. This formulation gives rise to Cragg's equator (7), where, with some change from Cragg's notation, we specify

$$Pr(y_{i1} = 1) = \phi(X_i \beta_1) \qquad\qquad (11)$$

$$Pr(y_{i1} = 0) = \phi(-X_i \beta_1),$$

where ϕ is the standard normal distribution function (Cragg uses $C(\cdot)$ for $\phi(\cdot)$).

For strictly positive second-stage quantity realizations, Cragg proposes two alternative formulations. Both are based on the specification

of the <u>conditional</u> densities for random variables Y_{i2} <u>given</u> that the activity is in fact undertaken.

The first formulation, described by Cragg's equation (9), is one where the conditional density for the realizations of the Y_{i2} is truncated-normal, with the truncation point at zero. Thus we have

$$g(y_{i2}|y_{i1}=1) = \phi(\frac{y_{i2}-X_i\beta_2}{\sigma})/\sigma\Phi(\frac{X_i\beta_2}{\sigma}), \qquad y_{i2} > 0 \qquad (12)$$

$$= 0 \qquad\qquad , \qquad \text{otherwise,}$$

Here ϕ and Φ are the standard normal density and distribution functions. With similar notational change, Cragg's equation (9), the (unconditional) likelihood of the positive realizations, can be written as

$$f(y_{i2}) = g(y_{i2}|y_{i1}=1)\Pr(y_{i1}=1) =$$

$$\phi(\frac{y_{i2}-X_i\beta_2}{\sigma})\Phi(X_i\beta_1)/\sigma\Phi(\frac{X_i\beta_2}{\sigma}) \qquad (13)$$

for $y_{i2} \gtrless 0$. Therefore, the likelihood function of the Cragg eqs. (7)-(9). model is

$$L = \prod_{i\in\Omega_0} \Phi(-X_i\beta_1) \prod_{i\in\Omega_1} \phi(\frac{y_{i2}-X_i\beta_2}{\sigma}) \Phi(X_i\beta_1)/\sigma\Phi(\frac{X_i\beta_2}{\sigma}) \qquad (14)$$

where Ω_0 is the index set for i such that $y_{i1} = 0$ and Ω_1 is the index set for $y_i = 1$. Written in log form,

$$\ell = \sum_{i\in\Omega_0} \ln(1-\Phi(X_i\beta_1)) + \sum_{i\in\Omega_1} \ln\phi(\frac{y_{i2}-X_i\beta_2}{\sigma}) + \ln\Phi(X_i\beta_1) - \ln\sigma - \ln\Phi(\frac{X_i\beta_2}{\sigma}). \quad (15)$$

In the form (15), it is straightforward to see that maximization of ℓ is fully equivalent to the two-stage maximization problem:

1) Probit estimation of the parameter vector β_1 via maximization of

$$\ell_1 = \sum_{i \in \Omega_0} (\ln(1-\Phi(X_i\beta_1))) + \sum_{i \in \Omega_1} \ln\Phi(X_i\beta_1); \tag{16}$$

2) Truncated-normal estimation of the parameters (β_2, σ) via maximization of

$$\ell_2 = \sum_{i \in \Omega_1} \ln\phi(\frac{y_{i2}-X_i\beta_2}{\sigma}) - \ln\sigma - \ln\Phi(\frac{X_i\beta_2}{\sigma}). \tag{17}$$

Because of the complexity of the log likelihood (15), estimation in this two-stage fashion is likely to be somewhat easier than attempting to maximize (15) with respect to the $(2k+1)$ parameters $(\beta_1, \beta_2, \sigma)$.

Cragg's second formulation again depends on the probit first-stage model, out the conditional density of the positive realizations is respecified. Instead of assuming that the conditional density of the positive realizations of y_{i2} is truncated-normal, the model is now formulated such that the logarithms of the y_{i2} are normal, i.e. conditional on $y_{i1}=0$, $\log(y_{i2}) \sim N(X_i\beta_2, \sigma^2)$. This is Cragg's equation (10). The condition& density for the $i \in \Omega_1$ is

$$h(y_{i2}|y_{i1}=1) = (y_{i2}\sigma)^{-1}\phi(\frac{\log(y_{i2})-X_i\beta_2}{\sigma}) \tag{18}$$

where the term $(y_{i2})^{-1}$ is the Jacobian of the transformation from y_{i2} to $\log(y_{i2})$. Therefore, the likelihood for the $i \in \Omega_1$, which is Cragg's equation (11), is

$$f(y_{i2}) = h(y_{i2}|y_{i1}=1) Pr(y_{i1}=1)$$

$$= (y_{i2}\sigma)^{-1} \phi(\frac{\log(y_{i2})-X_i\beta_2}{\sigma}) \Phi(X_i\beta_1) \qquad (19)$$

The likelihood function for the entire sample is

$$L = \prod_{i\epsilon\Omega_0} \Phi(-X_i\beta_1) \prod_{i\epsilon\Omega_1} (y_{i2}\sigma)^{-1} \phi(\frac{\log(y_{i2})-X_i\beta_2}{\sigma}) \Phi(X_i\beta_1) \qquad (20)$$

In log form,

$$\ell = \sum_{i\epsilon\Omega_0} \ln(1-\Phi(X_i\beta_1)) + \sum_{i\epsilon\Omega_1} \ln\phi(\frac{\log(y_{i2})-X_i\beta_2}{\sigma}) + \ln\Phi(X_i\beta_1)$$

$$-\ln y_{i2} - \ln\sigma \qquad (21)$$

As in the eqs. (7)-(9) model, the eqs. (7)-(11) model can be estimated in two stages:

1) Probit estimation of β_1 as above;

2) OLS estimation of (β_2,σ) using the log transform of the y_{i2} as dependent variables and X_i as the independent variables. This is perhaps surprising, but results because the terms in (21) involving (β_2,σ) are identical to those of the likelihood function of the familiar normal linear model.

Because of the simplicity of this two-stage approach, estimation in such a framework is obviously appealing. Duan, et. al. (1983) have proposed the Cragg (7)-(11) model to estimate medical expenditures: individuals either have or do not have medical expenses, and given that they have medical expenses, the conditional density of the expenditures is lognormal, $\log(Y_{i2}) \sim (X_i\beta_2,\sigma^2)$.

TRUNCATED-NORMAL ESTIMATION

As described above, estimation of the truncated-normal model is the relevant second step in estimating the Cragg (7)-(9) model where the positive observations are assumed to follow a truncated-from-below normal distribution. Although there are several variants of the truncated normal -- truncated-from-below, truncated-from-above, doubly-truncated; constant or nonconstant point(s) of truncation -- the discussion here will concentrate on the case most relevant to the present empirical work, viz. the truncated-from-below distribution where the point of truncation is constant across observations and is assumed to be zero. The results easily generalize, however, and for a discussion of the statistical properties of the truncated normal distribution in the most general case, the reader is referred to Johnson and Kotz (1970, pp. 81-87).

It should be noted that interest in the truncated normal should not be confined to the role it plays in the Cragg model. The distribution is useful in many empirical situations. Hurd (1979) notes that

(e)stimation based on only positive y's comes about very naturally in a number of kinds of studies. For example, in many labor supply studies one of the right-hand Variables, the wage rate, is only observed when the left-hand variable, labor supply, is positive. Imputing the unobserved wage rates causes a number of complications that can be avoided by discarding those observations for which labor supply is zero. Another example is a demand study where the price is not known unless a purchase is made. (Hurd, 1979, p. 248).

Furthermore, as we will see below, estimation of the truncated normal model on the nonlimit observations of a data set in which data on both limit and

nonlimit observations available, in conjunction with an informal test suggested by Olsen (1980), can give sane indication as to whether a Tobit model estimated on all observations is an appropriate specification.

For our purposes, the likelihood function of the truncated normal can be constructed as follows. We assume the existence of $T_1 + T_2$ realizations of random variables $Y_i \sim NID\ (X_i\beta, \sigma^2)$. However, for whatever reasons, only the positive realization of the Y_i are used in the analysis, these assumed to number T_1. Given these assumptions, the likelihood function is

$$L = \prod_{i=1}^{T_1} (\phi_i / \sigma \Phi_i),\qquad\qquad(22)$$

where ϕ_i is the standard normal density evaluated at $((y_i - X_i\beta)/\sigma)$ and Φ_i is the standard normal distribution function evaluated at $(X_i\beta/\sigma)$ which serves as the normalizing factor of the truncated density. The log-likelihood function (suppressing terms not depending on (β, σ)) is

$$\ell = \sum_{i=1}^{T_1} -.5((y_i - X_i\beta)/\sigma)^2 - \log \sigma - \log \Phi(\frac{X_i\beta}{\sigma})\qquad(23)$$

Estimation is by means of maximum likelihood. The first-order conditions for a maximum of ℓ are

$$\partial\ell/\partial\beta = \sum_{i=1}^{T_1} [\frac{-\phi_i}{\sigma\Phi_i} + (\frac{y_i - X_i\beta}{\sigma^2})] X_i' = 0$$

$$(24)$$

$$\partial\ell/\partial\sigma = \sum_{i=1}^{T_1} (\frac{\phi_i}{\Phi_i})(\frac{X_i\beta}{\sigma^2}) - \frac{1}{\sigma} + \frac{(y_i - X_i\beta)^2}{\sigma^3}.$$

The second derivatives are complicated and will not be presented here. Experience has demonstrated that the Berndt-Hall-Hall-Hauman first derivative approach for optimization works rather well.

In the case of the truncated-normal, as for almost all other limited dependent variable models, ordinary least squares estimation of the parameters (β, σ) yields biased and inconsistent estimates. However, in the case of the truncated normal, Olsen (1980) has shown how the OLS estimates can be used fruitfully to generate estimates of (β, σ) that, while inconsistent, can provide remarkably good approximations to the maximum likelihood estimates (our experience in other areas is consistent with Olsen's finding) and, as such, serve as excellent starting values for maximum likelihood estimation algorithms.

Olsen's method relies on a method of moments technique whereby the moments (specifically the mean and variance) of the empirical incomplete distribution, that of the positive y_i, are related to the moments of the complete distribution via formulae developed by Pearson and Lee (1908). Extending the Pearson-Lee methodology to the multiple regression case, Olsen demonstrates that the least squares slope coefficients differ from the true slope coefficients by a common factor, and he presents in tabular form the multiplicative correction factors needed to transform the OLS estimates of the slope, intercept, and standard error parameters (based or data from the incomplete distribution) to the corresponding complete distribution estimates. In practice, we have fitted polynominal functions of the third degree to Olsen's tabled data so that the transformations are facilitated.

Olsen also presents the multipliers for transforming the (mean/standard error) ratio estimated by OLS on the incomplete distribution

to the corresponding ratio of the complete distribution, (μ/σ). Olsen notes that $\Phi(\mu/\sigma)$, where Φ is the standard normal cumulative distribution, should give an idea of the expected ratio of nontruncated to total observations. Therefore, if one is considering Tobit estimation of the parameters of a censored distribution, it should hold that the $\Phi(\mu/\sigma)$ based on the (μ/σ) estimated using Olsen's method and treating the nonlimit observations as truncated normal should accord approximately with the ratio of noncensored to total observations. There is no formal test to assess how closely these should accord however. Olsen suggests that a disagreement here could well indicate that the Tobit is an inappropriate specification.

As is the case in the censored normal model discussed earlier, misspecification of the truncated-normal model has serious consequences for the consistency of maximum likelihood estimates. We describe briefly a general test for such misspecification.

Use of the Hausman (1978) specification test has become increasingly popular. Nelson (1981) has proposed a version of the test for misspecification of the censored-normal (Tobit) model. We here follow closely Nelson's development and adapt his test to the case where the model of interest is truncated-normal.

For the complete distribution where random variables $Y_i \sim N(X_i\beta, \sigma^2)$, the truncated-from-below normal density is defined by

$$f(y_i \mid y_i > 0) = f(y_i) = \phi((y_i - X_i\beta)/\sigma)/\sigma\Phi(X_i\beta/\sigma), \quad y_i > 0 \qquad (25)$$
$$= 0 \qquad\qquad , \text{ else,}$$

where ϕ and Φ are the standard normal density and distribution functions, and the point of truncation is assumed to be zero. As in the case of the censored-normal model, if the maintained hypotheses (e.g. errorless dependent variables, homoscedastic errors, normality) are violated, inconsistent estimates of the parameters (β, σ) will generally result if estimation is by maximum likelihood based on (25). This is, of course, analogous to the problems inherent when the censored-normal is misspecified.

The basic idea underlying the Nelson test is that there exist functions of the parameters of the model that under a large variety of circumstances are robust against misspecification of the underlying density. Such functions serve as the "consistent-inefficient" component of the Hausman test. The "inconsistent-efficient" component is the MLE of the model's parameters or (because of ML invariance properties (see Cox and Hinkley (1974, p. 287)), functions thereof estimated under the null hypothesis of no misspecification.

Because the censored- and truncated-normal densities are intimately related, we, like Nelson, use estimates of $E(T^{-1}X'Y)$ as the basis for the test. Here, X is the Txk matrix of independent variables, and Y is the Tx1 vector having typical element Y_i. Our development follows that of Nelson on pages 1327 and 1328 of his paper.

For the truncated-normal model as defined by (25) (see Johnson and Kotz (1970), pp. 81-87) we have:

$$E(Y_i) = X_i \beta + \sigma \phi(X_i \beta / \sigma) / \Phi(X_i \beta / \sigma) \tag{26}$$

$$E(Y_i^2) = (X_i\beta)^2 + \sigma^2 + (X_i\beta)\sigma\varphi(X_i\beta/\sigma)/\Phi(X_i\beta/\sigma) \tag{27}$$

$$E_{XY} = E(T^{-1}X'Y) = T^{-1}(X'X\beta + \sigma X'\lambda) \tag{28}$$

where λ is the Tx1 vector with typical element (φ_i/Φ_i). The method of moments estimator of $E(T^{-1}X'Y)$ is $T^{-1}X'y$. The limiting variance of $T^{.5}(T^{-1}X'y)$ is

$$V_1 = T^{-1}X'V_Y X, \tag{29}$$

where V_Y is the TxT diagonal matrix with typical element $(E(Y_i^2) - E^2(Y_i))$ as defined in (26) and (27).

The efficient estimator of $E(T^{-1}X'Y)$, denoted $\hat{E}_{XY}$, is obtained by evaluating (28) at $(\hat{\beta}, \hat{\sigma})$ ("^" signifies a MLE). The limiting variance of $\hat{E}_{XY}$ is obtained via the analog to the approximation of Nelson's equation. (3.9) and is

$$V_0 = T^{-2}[\ X'X \mid X'\lambda\ ]C(\hat{\beta}, \hat{\sigma})[\frac{X'X}{\lambda'X}] \tag{30}$$

evaluated at $(\hat{\beta}, \hat{\sigma})$ It is estimated as $\hat{V}_0$ by evaluating λ at $(\hat{\beta}, \hat{\sigma})$. $C(\cdot) = TI(\cdot)^{-1}$ where I is the estimated information matrix.

The test statistic is

$$m = T[T^{-1}X'y - \hat{E}_{XY}]'[\hat{V}_1 - \hat{V}_0]^{-1}[T^{-1}X'y - \hat{E}_{XY}], \tag{31}$$

where $\hat{V}_1$ is (29) evaluated at $(\hat{\beta}, \hat{\sigma})$. Under the null hypothesis of no misspecification, $m \sim \chi^2_{(k)}$.

HECKMAN'S APPROACH: SAMPLE SELECTION BIAS

Currently, the most prevalent limited dependent variable estimation technique is the sample selection bias model, attributable largely to Heckman (1976, 1979). The model has a number of applications (see Heckman's 1976 article in particular), and is quite easy to estimate. Because it is so well-known, we will only provide a sketch of the details. The following section, which contrasts and compares the Tobit, Cragg, and Heckman models, sheds some more light on subtleties of Heckman's formulation.

Heckman considers the following two-equation model:

$$y_{i1} = X_i \beta_1 + \epsilon_{i1} \tag{32}$$

$$y_{i2}^* = X_i \beta_2 + \epsilon_{i2} \tag{33}$$

It is assumed that ϵ_i and ϵ_{i2} are distributed joint normal, with marginal densities $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ respectively, and covariance σ_{12}. It can be further assumed that the realizations y_{i2}^* are unobserved. However, discrete sign indicators y_{i2} are available and are mapped as

$$y_{i2} = 1, \; y_{i2}^* > 0 \tag{34}$$
$$\quad = 0, \; y_{i2}^* \leq 0.$$

In Heckman's model, the realizations y_{i1} are available to the analyst only when $y_{i2}^* >$ 0, i.e. when $y_{i2} = 1$.

A concrete example is where (32) is a model determining market wage rate (or log(wage rate)) by a linear function of X_i and random error and where (33) is a model determining hours of labor supplied in the market. It is assumed that either hours of labor supplied or a discrete binary indicator of whether or not any hours were supplied is available for all observations. However, because market wage rates are only observed for

individuals for whom the market wage rate exceeds the reservation wage at zero hours, data on the y_{i1} are available only when $y_{i2}^* > 0$ $(y_{i2} = 1)$.

Heckman then considers the expectation $E(Y_{i1}|y_{i2} = 1)$, which can be written as

$$E(Y_{i1}|y_{i2} = 1) = X_i\beta_1 + E(\varepsilon_{i1}|y_{i2} = 1). \tag{35}$$

If one considers least-squares estimation of (35), the question is: Are the estimates of β_1 consistent when y_{i1} is regressed on those X_i for which $y_{i2}=1$? Basically the issue is whether the expectation $E(\varepsilon_{i1}|y_{i2} = 1)$ is null. In general, and thus at the core of the sample selection bias problem, the answer is "no". Based on well-known formulae, it holds that

$$E(\varepsilon_{i1}|y_{i2} = 1) = \sigma_{12}\phi_i/\sigma_2(1-\Phi_i), \tag{36}$$

where ϕ_i is the standard normal density evaluated at $(X_i\beta_2/\sigma_2)$ and Φ_i is the distribution function evaluated at the same point. Because σ_{12} is in general nonzero and since ϕ_i, $(1-\Phi_i)$, and σ_2 are all positive, then least squares estimation of (35) will be based on an expectations function with nonnull disturbance expectation, and will therefore yield inconsistent estimates of β_1.

Heckman's suggested procedure in this situation is as follows. Estimate on the entire sample a probit model for the discrete indicator representation of the model (33). This yields a consistent estimate of the parameter vector (β_2/σ_2) from which consistent estimates of $\lambda_i = \phi_i/(1-\Phi_i)$ are constructed. Form the Tx(k+1) matrix Z = $[X|\Lambda]$, where A is a Tx1 vector with typical element λ_i, and regress y_i on$[x_i, \lambda_i]$. This procedure yields consistent estimates of the parameters β_1 and (σ_{12}/σ_2), having

effectively solved the omitted variables problem by using a consistent estimate of $E(\varepsilon_{i1}|y_{i2} = 1)$ as a regressor.

In the context of participation models, one could define y_{i2} as some latent index of the desire to participate. Given that this index is greater than some threshold level, participation results, its magnitude determined by the realization y_{i1}. The translation of the participation model into Heckman's framework is not straightforward, however. For nonparticipants, we <u>observe</u> zero hours of participation rather than not observing the amount. It is therefore difficult to interpret the meaning of the realized, but unobserved, y_{i1} for nonparticipants. We turn in the next section to a more detailed analysis of such subtleties.

TOBIN, CRAGG, AND HECKMAN: A DIGRESSION

As there are some similarities between and among the models described above and identified for expositional parsimony as the models of Tobin, Cragg, and Heckman, it is probably appropriate to summarize their similarities and differences and in so doing to elucidate the circumstances in which each model is more or less appropriate. (The discussion of Cragg's model here is the Cragg (7)-(9), i.e., probit/truncated-normal, model as that version is most similar to the others discussed here.)

First note that the Tobit model is a restricted version of both the Cragg and the Heckman models. The reason for this is purely mechanical, however, and should not be taken to imply that the Cragg and Heckman models are in general identical. As we will see below, these models are structurally quite different.

To see that the Cragg model reduces to the Tobit, the Cragg log-likelihood function can be written (following Lin and Schmidt (LS)

(1984)) as

$$\ell = \sum_{i \in \Omega_0} \ln\Phi(-X_i \beta_1) + \sum_{i \in \Omega_1} \left[\ln\Phi(X_i \beta_1) - \ln\phi(\frac{X_i \beta_2}{\sigma}) - \right.$$

(37)

$$\left. (1/2)\ln(2\Pi\sigma^2) - (1/2\sigma^2)(y_i - X_i \beta_2)^2 \right]$$

If the restriction $\beta_1 = \beta_2/\sigma$ is imposed, then the first two terms in the square brackets cancel and (37) is easily seen to be identical to (6) with β in (6) replaced by β_2 from (37). The upshot of such parameter restrictions is hardly trivial, however. As specified, and discussed briefly earlier, the Tobit model is fairly restrictive in its behavioral implications, as the parameter vector that governs the probability of observing an above-threshold realization of the dependent variable is the same as that governing the quantity realization of the dependent variable given that it is above the threshold. Owing to the implications of such restrictions, LS have concluded that "...the Tobit model is typically used with more faith than it warrants..." and have developed a test (which we discuss below) for the appropriateness of the $\beta_1 = \beta_2/\sigma$ restriction of the Cragg model that is implied by the Tobit specification. The following excerpt from LS provides a particularly cogent summary description of the appropriateness of the restricted (Tobit) versus the unrestricted versions of the Cragg model Lin and Schmidt, 1981, pp. 174,5):

(I)n the Tobit model any variable which increases the probability of a non-zero value must also increase the mean of the positive values; a positive element of β means that an increase in the corresponding variable (element of X_i) increases both $Pr(y_i > 0)$ and $E(y_i | y_i > 0)$.   This is not always reasonable. As an example,

consider a hypothetical sample of buildings, and suppose that we wish to analyze the dependent variable "loss due to fire," during some time period. Since this is often zero but otherwise positive, the Tobit model might be an obvious choice. However, it is not hard to imagine that newer (and more valuable) buildings might be less likely to have fires, but might have greater average losses when a fire did occur. The Tobit model can not accommodate this possibility.

Another problem with the Tobit model is that it links the shape of the distribution of the positive observations and the probability of a positive observation. For rare events (like fires), the shape of the distribution of the positive observations would have to resemble the extreme upper tail of a normal, which would imply a continuous and faster than exponential decline in density as one moved away from zero. Conversely, when zero occurs less than half of the time, the Tobit model necessarily implies a non-zero mode for the non-zero observations.

Cragg's model avoids both of the above problems with the Tobit model. A reasonably strong case can be made for it as a general alternative to the Tobit model, for analysis of data sets to which Tobit is typically applied--namely, data sets in which zero is a common (and meaningful) value of the dependent variable and the non-zero observations are all positive. The distribution of such a dependent variable is characterized by the probability that it equals zero and by the (conditional) distribution of the

positive observations, both of which Cragg's model parameterizes

in a general way.

As mentioned above, a formal test of the validity of the restrictions on the Cragg (7)-(9) model, such that the restrictions imply the Tobit specification, has been proposed by Lin and Schmidt. Their observation is that since the Tobit model can be viewed as a restricted Cragg model, a straightforward test for the validity of the restrictions (that furthermore circumvents the need to estimate both the restricted and unrestricted forms of the model) is a Lagrange multiplier test.[4] The simplicity of the test is extremely attractive. Based on the results of the Tobit estimation, the Lagrange multiplier statistic is calculated, and, using a χ^2 test, the validity of the restrictions is tested. If the test indicates rejection of the null hypothesis that the restrictions hold, then the Cragg (7)-(9) maximum likelihood estimates can be obtained via probit and truncated-normal estimation. Should the test fail to reject the null hypothesis, however, the analyst is then spared the effort and expense of estimating the unrestricted Cragg model. However, it is not always the case that the Tobit is computationally less burdensome than the Cragg alternative in which instances the appeal of the LM test is somewhat diminished. We propose here a Wald test for the Tobit parameter restrictions considered by Lin and Schmidt that might be considered attractive when the Cragg model is computationally preferred to the Tobit

Some recent experience has found the Tobit model in several empirical applications to be consistently rejected in favor of Cragg's alternative based on the Lagrange multiplier test criterion. Should these specific results generalize, efficient estimates of the parameters of the Cragg model are likely to be desired by researchers otherwise contemplating the

Tobit approach. The appeal of the Wald test, then, is that since only the estimates of the parameters of the unrestricted model are required, maximum likelihood estimates of the Cragg model are available immediately and the Tobit model need not be estimated. Regardless of whether the Wald test suggests rejection of the Tobit restrictions, a model with a necessarily higher likelihood than the Tobit will have been estimated in the first instance. It should be noted that the Wald test is in a sense a mirror image of the Lagrange multiplier test in that the former relies solely on. ML estimation of the unrestricted (here, Cragg) model while the latter is based exclusively on the ML parameter estimates of the restricted (here, Tobit) specification. The two test statistics, however, have the same asymptotic distribution (see Rao (1965) pp. 347-352).

Following Rao (1965) and Amemiya (1983), the Wald test statistic is

$$W = h'[HJ^{-1}H']^{-1}h, \tag{38}$$

where $J=\hat{J}(e)$ is the estimated information matrix, $h=\hat{h}(e)$ is the (kx1) vector of nonlinear restrictions on the parameters of the form $n(\hat{\theta})=0$, and $H=\hat{H}(\theta)\partial n/\partial\theta'$ is a (kxq) matrix of partial derivatives. All evaluations are at $\hat{\theta}_{ML}$ which is the ML estimate of the parameters of the unrestricted (here, Cragg) model. Under appropriate conditions (see Rao), W is distributed asymptotically central χ^2 with k degrees of freedom under the null hypothesis that $\beta_1=\beta_2/\sigma.$

In the Cragg specification, $\theta=(\beta'\beta_2',\sigma)'$ and q=(2k+1). The, parameter restrictions to be tested can be written as

$$h = \sigma\beta_1-\beta_2. \tag{39}$$

Given this form for h, it follows that

$$H = [\sigma I_k|-I_k|\beta_1]. \tag{40}$$

J^{-1} can be estimated by the ML covariance matrix of the parameters of the

Cragg model. Because of the structure, J, and therefore J^{-1}, are block diagonal, viz.

$$J^{-1} = \begin{bmatrix} J^{11} & 0_{k \times k} & 0_{k \times 1} \\ 0_{k \times k} & J^{22} & J^{23} \\ 0_{1 \times k} & J^{32} & J^{33} \end{bmatrix} = \begin{bmatrix} A & 0' \\ 0 & B \end{bmatrix} \tag{41}$$

so that the submatrices A and B can be estimated separately *as* the covariance matrices of the probit parameters β_1 (A) and the truncated-normal parameters (β_2, σ) (B).

Thus, W can be calculated by evaluating (39), (40), and (41) at $\hat{\epsilon}_{ML}$ and using the formula (38). As mentioned above, under the hull hypothesis that the restrictions hold, W is distributed asympotically central x^2 with k degrees of freedom. Furthermore, W has the same asymptotic distribution as the Lagrange multiplier test statistic proposed by LS, so preference for one test statistic over the other will likely depend largely on the relative ease of implementation.

Turning now to Heckman's formulation, his two-equation model is seen to reduce to the Tobit model as follows. Recall that the model can be written (with notational changes) as

$$y_{i1} = X_i \beta_1 + \epsilon_{i1}$$

$$\tag{42}$$

$$y_{i2}^* = X_i \beta_2 + \epsilon_{i2}.$$

Y_{i2}^* is a latent variable, however, and only a discrete (0,1) sign indicator of its realization y_{i2} is available. y_{i1} is observed <u>only</u> when $y_{i2} = 1$. Letting $\beta_1 = \beta_2$ and $\epsilon_{i1} = \epsilon_{i2}$ (i.e. the error structure is univariate rather than bivariate), then the Heckman model is the standard Tobit model. The

logic is that when these restrictions are imposed in the Heckman two-equation model, the remaining single equation plays both the censoring and the determination-of-intensity roles. Since the censoring occurs as a result of a non-positive realization of the random variable Y_{i2}^*, the Tobit requirement that the quantity or intensity realization be confined to the nonnegative orthant is automatically satisfied when the restriction $y_{i1} = y_{i2}^*$ (i.e. $\beta_1 = \beta_2$, $\varepsilon_{i1} = \varepsilon_{i2}$) is imposed. In general, however, the Heckman two-equation framework is not specifically designed to model situations where realizations of the dependent variable of interest are necessarily nonnegative and are recorded for all individuals/observations, and where $Pr(y_i = 0) > 0$. Heckman's formulation has $y_i \neq 0$ except on a set of measure zero. We turn now to an explanation of the fundamental differences between the Heckman two-equation formulation and the two versions-of-interest of the Cragg model.

The two-equation Heckman model describes two phenomenon, Y_{i1} and Y_{i2}^*, that are marginally distribute, respectively, as $NID(X_i \beta_1,) \sigma^2$ and $NID(X_i \beta_2, \sigma_2^2)$ (σ_2^2 is usually restricted $= 1$ for normalization when only the sign of y_{i2}^* is observed). The joint distribution is bivariate $NID(X_i \beta_1, X_i \beta_2, \sigma_1^2, \sigma_1^2, \rho)$, where ρ is the correlation of $(\varepsilon_{i1}, \varepsilon_{i2})$, $(\sigma_{12}/\sigma_1\sigma_2)$, which is in general nonzero. The important point is that these marginal and joint distributions are <u>unconditional</u>. That is, for <u>all</u> i, there exist realizations (y_{i1}, y_{i2}^*) although the realizations y_{i1} for some i will be unavailable to the researcher. Casting the problem concretely in the area where Heckman's model has been most fruitfully applied, labor economics, sheds further light on the subtleties of his model. Here we define $y_{i1} = \log(W_i)$ and $y_{i2}^* = \log(H_i + 1)$, where W_i is wage earned in market work and H_i is hours of market work. Thus, y_{i2}^* is positive only if market hours are

positive. It is posited that the expected values of both Y_{i1} and Y_{i2} are linear functions of personal characteristics and other variables so that the two-equation model results. However, because we only observe the market 'wage for those individuals actually participating in market work (those for whom $H_i>0$), some subset of observations will not have data on the y_{i1}. There is a market wage determined for nonparticipants; whether or not such individuals have knowledge of their market wages is immaterial. The relevant analytical fact is that such data are unavailable to the researcher.

In this labor supply framework, it is apparent why the estimation techniques developed for the two-equation Heckman model and discussed earlier in this chapter have such appeal. The more immediate concern, of course, is whet her such techniques are in fact appropriate to the estimation requirements of the present analysis. In a nutshell, Heckman's model is one where there are two equations of interest, both holding for all i unconditionally, and where (except when restricted so as to be identical to a Tobit model) the probability of observing realizations of the dependent variable equal zero is zero. Does such a formulation capture the essence of the "corner solution" problems of the participation decision?

It seems rather artificial to cast the recreation participation functions in such a framework. It is not generally the case with the generation of participation data that we can posit the existence of some latent variable such that data for the participation measure(s) of interest are only available given a positive realization of the latent variable. Rather, the processes of interest here are represented more typically by data that indicate the realizations of participation decisions for all

individuals, even though these realizations are quite frequently on the boundary of the consumption set.   We turn now to a discussion of how the Cragg models differ in substance from the Heckman two-equation setup and argue that the Cragg formulations are more suited than Heckman's model to the nature of a subset of our estimation requirements.

Although like the Heckman formulation in being a "two-model" specification, the fundamental point of departure for the Cragg technique is that one of the two models is formulated in terms of <u>conditional</u> expectations.   The conditions on which the expectations are taken are, as described above, the outcomes of unconditional models, which are generally stated as binary representations of latent random variables.   Thus, in the context of recreation participation, there is an unconditional model defined for all individuals determining the binary outcome (participate, don't participate).   <u>Conditional on</u> a "participate" outcome, the quantity of participation is determined either by a lognormal or truncated-normal model.   The unconditional likelihood for a representative participant is then

density (participation given participate) *Pr (participate), (43)

which is equation (13) as specified earlier. There is no density of the quantity of participation defined for nonparticipants, unlike Heckman's formulation that defines such a density for all individuals.

Deaton and Irish (DI) (1984) , in an. independent line of investigation, have cast the Cragg (7)-(9) model in a two-equation Heckman formulation. They indicate that a positive observation on the quantity measure of interest is made when, in the notation used earlier, both Y_{i1} and Y^*_{i2} are realized as positive, else a zero or a nonparticipation results. In two cases, DI specify

$$y_{i1} = X_i\beta_1 + \varepsilon_{i1}$$

(44)

$$y_i^* = X_i\beta_2 + \varepsilon_{i2}$$

cast thusly, the Cragg (7)-(9) model <u>can</u> be viewed as a Heckman two-equation model, but with a restriction imposed that is absent in Heckman's formulations. That is DI seem to have ignored one aspect of the Cragg model that is key in differentiating it from Heckman's specification, viz. that $y_{i2}^* > 0$ is both a necessary and sufficient condition for a positive realization of y_{i1} to result. That is, $Pr(y_{i1} > 0 | y_{i2}^* > 0) = 1$, $Pr(y_{i1} = 0 | y_{i2}^* < 0) = 1$. When, and only when the first hurdle is traversed is there a positive amount of the activity undertaken. So DI's statement that positive realizations of <u>both</u> variables determines whether   is observed positive is somewhat misleading in that a positive realization of either suffices to assume the positivity of the other. Neither of Cragg's specifications, then, is really in the spirit of the model proposed by Heckman except, of course, when both the Cragg (7)-(9) model and the Heckman two-equation formulation are restricted such that the Tobit specification results.

Owing to the subtleties of the arguments, it is likely that the above discussion has provided somewhat less than a total clarification of all the relevant issues. Some of these shortcomings are due to the fact that even Central participants in the academic debates appear still unconvinced about the nature of the differences among the estimation techniques. For example, as noted earlier Duan and coauthors (1983) have used the Cragg (7)-(11) estimation technique to model individuals' medical expenditures. The expenditure decision, in the spirit of Cragg's specification, is statistically modeled as two separate processes. Model one determines the

binary outcome of whether or not any expenditures will occur, and model two determines the amount of expenditure (positive by definition) that results conditional on there being some expenditure. In this paper, Duan and coauthors assert that the covariance between the error terms of the two models is irrelevant insofar as construction of the likelihood function is concerned.

Recently, however, Hay and Olsen (1984) have questioned the Duan and coauthors method, stating that this approach "requires some fairly unusual assumptions on the model joint error distribution and functional form (p. 279)." Moreover, Hay and Olsen go on to claim that the Duan and coauthors formulation "can be interpreted as being nested in the more general sample selection models (p.279)." Duan and coauthors respond that Hay and Olsen "are incorrect in claiming that our models are nested within the sample selection model." and that "the conditional specification in the multi-part (i.e., Duan and coauthors) model is preferable to the unconditional specification in the selection model for modeling actual (v. potential) outcomes (p.283)."

As we argued earlier, the sample selection or Heckman approach Is particularly fruitful when analyzing phenomena such as labor market participation. Quoting Duan and coauthors:

> For certain empirical problems such as labor force participation, the primary goal might be to predict the potential outcome instead of the actual outcome; therefore, an unconditional specification such as the sample selection models might be preferable. For the present application, however, the goal is to predict the actual expense, not the potential expense; therefore, the unconditional equation... is of no

direct interest, and the preference for the unconditional specification in the other empirical problems does not apply to the present application. (p. 286).

In any event, this discussion demonstrates that there still exists some confusion on these points in the published literature. We have attempted to be as thorough as time and space permit in hope of emphasizing one extremely important message. That is, it is essential that the researcher be intimately familiar with the behavioral and statistical structure of the models of interest in order to avoid being swallowed by the slippery quicksand we have described. The nature of participation measures as conditional or unconditional and the interpretation of any latent variables in the model must be quite clear before the correct estimation technique can be selected. When, and only when, such issues are in order is it possible to make sense of the estimated obtained and their relevance to benefit estimation.

It seems that the logic of the participation decisions of interest in this study is better captured in terms of Cragg's specifications than in the Heckman two-equation model although this question is obviously still open to informed debate. The specification of the magnitude-of-participation model as a conditional model is, however, intuitively plausible, and Cragg's formulations provide a natural vehicle for translating such intuitive plausibility into an econometric framework.

POISSON-DISTRIBUTED PARTICIPATION DAYS

In modeling event counts (non-negative integer data) over some time interval $(t, t+dt)$, the Poisson distribution is commonly used. Here, a random variable Y_i follows the probability law

$$\Pr(Y_i = y) = \exp(-\lambda_i)\lambda_i^y/y! , \quad y\varepsilon\{0,1,2,\ldots\}$$
$$= 0 \qquad\qquad , \text{ else} \tag{45}$$

with $E(Y_i) = \text{Var}(Y_i) = \lambda_i$.

It happens that there exist recreational participation data of interest that are recorded as nonnegative integers, most obviously as counts of days of participation. For any individual, such measures can, over a time interval (t, t+dt), say one year, assume only integer values in $\{0,1,2,\ldots,365\}$. Because of the paucity of observations likely to be found at the upper (365 day) limit, we ignore the fact that these measures obey upper bounds and concentrate instead on the complications presented by the large number of individuals who in a typical random sample of the population report zero days of participation in the relevant categories.

Analogous to the familiar normal distribution where for econometric work one typically specifies $\mu_i = X_i\beta$, the λ_i parameter of the Poisson distribution can be reparameterized to admit the influence of covariates. Since for all i, $\lambda_i > 0$, a straightforward approach is to assume $\lambda_i = \exp(X_i\beta)$ and to estimate β by maximum likelihood (see Hausman, Hall, Griliches (1984)) Hausman, Ostro, Wise (1983), Portney and Mullahy (1984)). This is the approach adopted here for modeling the participation-days outcomes.

One drawback of the Poisson model is the restriction that $E(Y_i) = \text{Var}(Y_i)$. Should this restriction not in fact characterize the data, the maximum likelihood estimates of the covariance matrix of $\hat{\beta}$ based on minus the inverse of the estimated Hessian will be inconsistent and t-tests based thereon would be misleading. Hausman, Ostro and Wise circumvent this restriction by allowing for an overdispersion parameter. A different approach is used here, using an estimator of the covariance matrix that is

robust against departures from the mean=variance restriction, this procedure is described below.

Given T independent observations, the log-likelihood function of the Poisson participation model can be written as

$$\ell = \sum_i -\exp(X_i \beta) + y_i X_i \beta + C, \tag{46}$$

where $\exp(X_i \beta) = \lambda_i$, y_i is the observed participation day count, and C does not depend on B. It can be shown that ℓ is concave in B. The first-order conditions for the maximization of ℓ are

$$\partial \ell / \partial \beta = \sum_i -\exp(X_i \beta) X_i' + y_i X_i' = 0 \tag{47}$$

with the maximum guaranteed by the condition

$$\partial^2 \ell / \partial \beta \partial \beta' = \sum_i -(X_i' X_i) \exp(X_i \beta) \tag{48}$$

negative definite.

The maximum likelihood estimates of β obtained by maximizing (46) are consistent, but the estimate of the covariance matrix of $\hat{\beta}_{ML}$ using $[-\partial^2 \ell / \partial \beta \partial \beta']^{-1}$ evaluated at $\hat{\beta}_{ML}$ will be inconsistent if the data are not in fact generated by the specified Poisson distribution.

This is most easily seen as follows. Note that the model can be equivalently cast as a nonlinear least squares regression, the i-th observation being

$$y_i = E(Y_i) + \varepsilon_i$$

$$\tag{49}$$

$$= \exp(X_i \beta) + \varepsilon_i$$

with $E(\varepsilon_i) = 0$. Clearly, $\text{var}(\varepsilon_i) = \text{var} : \cdot \cdot \exp(X_i \beta)$, so that the ε_i are

heteroscedastic. If nonlinear weighted least squares is used with the weights $\exp(-X_i\beta)$ formed using consistent estimates of β, and if the data are in fact Poisson as specified, the maximum likelihood consistent estimates of β and $\text{cov}(\hat{\beta})$ will obtain. (The consistency of $\hat{\beta}_{ML}$ for β does not depend on the weighting scheme.) However, if the data is not Poisson-distributed, the estimate of $\text{cov}(\hat{\beta})$ obtained in this manner will be inconsistent and asymptotic t-tests based thereon will be misleading. The case is fully analogous to the estimation of the heteroscedastic linear model which yields inconsistent covariance estimates (and, therefore, t-statistics) if the heteroscedastic nature of the error structure is either ignored or incorrectly specified.

Royall (1984) has demonstrated a method whereby estimates of $\text{cov}(\hat{\beta})$ robust against misspecification of the underlying distribution of the data can be obtained for various distributions, including the Poisson, when $[-\partial^2\ell/\partial\beta\partial\beta']^{-1}$ evaluated at $\hat{\beta}_{ML}$ fails to yield a consistent estimate of $\text{cov}(\hat{\beta})$. Denoting $I(\beta)$ as $[-\partial^2\ell/\partial\beta\partial\beta']$, Royall's suggestion is to estimate $\text{cov}(\hat{\beta})$ as

$$I(\beta)^{-1}[\Sigma_i(\partial\ell_i/\partial\beta)(\partial\ell_i/\partial\beta)']I(\beta)^{-1} \tag{50}$$

where ℓ_i is the i-th observation's contribution to the log-likelihood function and where all relevant evaluations in (50) are at $\hat{\beta}_{ML}$.

GEOMETRIC-DISTRIBUTED PARTICIPATION DAYS

One alternative to the Poisson model for the modeling of count data is the geometric distribution. Though seemingly not as often used by econometricians as the Poisson, the geometric is a logical choice should an

alternative to the Poisson be desired. Furthermore, the basic geometric specification does not suffer from the mean=variance restriction that is implied in the basic Poisson model. As will be seen below, the variance of a geometric-distributed discrete random variable is greater than its mean, although the fact that the variance depends on the mean limits somewhat the flexibility of the distribution.

Our description of the properties of the geometric distribution follows that of Johnson and Kotz (1969). First, it should be noted that the geometric is a special case of the negative binomial. Discussion is confined here to the geometric because it is computationally far more straightforward than is the general negative binomial. The geometric distribution is defined as follows:

$$Pr(X=k) = P^k(1+P)^{-(k+1)}, \quad k = 0,1,2...$$ (51)

$$= 0 \quad , \text{ else}$$

with $P>0$. It holds that $E(X) = P$ and $Var(X) = P(1+P)$. As in the econometric specification of the Poisson model considered earlier, one allows the P to vary across observations as P_i, and again $P_i = \exp(X_i\beta)$ is a sensible parameterization due to the required positivity of the P_i.

Given this, the likelihood function for T independent observations car. be written as

$$L = \prod_{i=1}^{T} \exp(k_i X_i \beta)(1 + \exp(X_i\beta))^{-(k_i+1)}$$ (52)

with log equal to

$$\ell = \sum_{i=1}^{T} k_i x_i \beta - (k_i+1) \log(1 + \exp(x_i\beta))$$ (53)

where k_i is the observed count for the i-th observation. The ML estimate $\hat{\beta}$ satisfies

$$\partial \ell / \partial \beta = \sum_{i=1}^{T} [k_i - (k_i + 1)\exp(X_i \beta)/(1 + \exp(X_i \beta))]X_i' = 0 \tag{54}$$

The Hessian is

$$H = \partial^2 \ell / \partial \beta \partial \beta' = \sum_{i=1}^{T} -(k_i + 1)[\exp(X_i \beta)/(1 + \exp(X_i \beta))^2]X_i' X_i,$$

which is seen by inspection to be negative definite. Because it is a fairly uncluttered expression, estimation and inference can proceed using -H as an estimate of the information matrix and $(-H)^{-1}$ as an estimate of the covariance matrix. (It might also be noted that (55) bears a strong resemblance to the Hessian of the well-known binary logit model.[5] Unfortunately, much like the Poisson specification, the covariance estimate thus obtained is not robust to departures from the data being in fact geometric. However, the methods proposed by Royall (1984) and described for the Poisson model can be used for the geometric distribution also. As the development is identical, the details are omitted.

MULTINOMIAL-DISTRIBUTED PARTICIPATION DAYS

One type of micro data of particular interest in recreation economics is of the following nature. We observe over the course of sane fixed time period (say one year) the number of times (say days) that an individual participates in (k-1) mutually exclusive recreation activities and, therefore, the number of days on which no recreational activity was undertaken which can be viewed as the k-th activity. To be concrete, the annual recreation profile for some individual who has in his/her recreation

possibility Set three activities (fishing day (=F), boating day (=B), swimming day (=S)) and nonparticipation (=N=365-F-B-S) days might look like

F = 12

B = 17

S = 0

N = 336

We also presume that the profiles of M individuals are observed.

In the analysis of such data, it is helpful to make two (fairly strong) assumptions:

(1) the data characterizing the individuals, i.e, the independent variables, are invariant over the fixed time period. That is, the characteristics of individual i, X_i, are representative of i for the entire year;

and

(2) the decision to participate in any one activity -- including nonparticipation -- on any given day depends neither on what. activities have been undertaken on the previous days nor or expectations of recreational participation in future days. That is, the daily decisions are (statistically, at best) independent.

Note that both these assumptions are more or less questionable with. (2) perhaps being the more restrictive assumption. However, we proceed under the constraints that these assumptions impose.

Given observations on the type of recreational profiles described above, and the assumptions there set forth, it is appropriate to view the data characterizing the recreation participation of individuals as realizations of multinomial random variables (see Morey (1981) for a

related discussion). From discrete statistical theory, the multinomial distribution of a random variable $\underset{\sim}{Y}$ with parameters $(T; P_1, \ldots, P_k)$ can be written

$$Pr(\underset{\sim}{Y} = \underset{\sim}{y}) = T! \prod_{j=1}^{k} (P_j^{t_j}/t_j!), \tag{56}$$

where T is the number of trials (here days), the t_j are the number of occurrences of the j-th outcome, and P_j are the probabilities that the j-th outcome will occur on a single trial.

To extend the statistical model to the recreation participation measure, we consider each daily participation decision (where "participation" now refers to participation in nonrecreational activities also) as one trial from a multinomial distribution with individual-specific parameter vector for the m-th individual $(T_m; P_{1_m}, \ldots, P_{k_m})$. Assuming $T_m = T'_m = T$ for all m, m', we henceforth drop the subscripts on the T parameters. The yearly profile, then, is the 365 (by assumption independent) daily trials for each individual. The econometric objective is the estimation of the P_{j_m}, i.e. estimation of the probabilities of engaging in the k possible activities on a given day.

For computational simplicity, we proceed as follows. A logistic distribution for the daily outcome probabilities is assumed. Thus, the probability that the outcome is Z on any trial is

$$P_{Z_m} = \exp(X_m \beta_Z) / \sum_{j \in \Omega} \exp(X_m \beta_j) \tag{57}$$

for $Z \in \Omega = \{F, B, S, N\}$. The logistic distribution assures that for all m

the multinomial requirement $(\Sigma_{j\epsilon\Omega} P_{j_m} = 1)$ is met.

Since the probabilities (57) are unique only up to a difference in parameter vectors $(\beta_j - \beta_{j'})$, some normalization is required. The normalization most convenient and easily interpreted is $\beta_N = 0$, so that β_F, β_B, and β_S are interpreted as differences between the respective activity parameter vectors and the nonparticipation parameter vector.

The objective, then, is estimation of the parameter vectors β_F, β_B, and β_S. This is, of course, fully analogous to the widely-used multinomial logit model where a single outcome from a set of mutually exclusive outcomes is considered. In fact, that case is merely a special case of the present exposition for which $T_m = 1$ for all m.

Estimation is by means of maximum likelihood. Assuming the existence of N independent profile draws from the population, the likelihood of the data as a function of the parameters is

$$L(\beta) = \prod_{m=1}^{M} Pr(\underset{\sim}{y}_m = \underset{\sim}{y}) = \prod_{m=1}^{M} T! \prod_{j\epsilon\Omega} (P_{j_m}^{t_{j_m}}/t_{j_m}!) \tag{58}$$

where the P_{j_m} are as defined in (57) and where Ω is the choice index set. In log form,

$$\ell(\beta) = \sum_{m=1}^{M} \sum_{j\epsilon\Omega} t_{j_m} \log P_{j_m} + C, \tag{59}$$

where C is a constant not depending on β. Given the assumed logistic probabilities, we have

$$\ell(\beta) = \sum_{m=1}^{M} \sum_{j\epsilon\Omega} t_{j_m} [X_m\beta_j - \log (\sum_{k\epsilon\Omega} \exp (X_m\beta_k))] + C. \tag{60}$$

Maximizing (45) is simpler than maximizing (43), and can be accomplished with only a slight modification of most existing (single-trial) multinomial logit programs. [6]

GROUPED OR INTERVAL DATA - ESTIMATION UNDER THE NORMALITY ASSUMPTION

There are often institutional or other constraints in the sampling or data-recording processes that have the effect of generating inexact data for research purposes. A common case and one that is of immediate relevance insofar as the present empirical investigations are concerned is the situation where continuous measures of interest, such as the amount of time spent participating in sane recreational activity, are cast in the recorded micro data as grouped or interval data. We discussed above strategies that might be considered when the outcomes are recorded as "number of days" or "number of times," i.e. where the data can be viewed as realizations of discrete statistical processes rather than as discrete/ integer codings of fundamentally continuous processes. In this section we concern ourselves with the situations where the underlying processes are best viewed as continuous phenomena but where the vagaries of either the sampling or data-coding procedures are such that only a finite number of intervals which the continuous measure is defined are determined and the only data available to the analyst are indicators of the interval bounds in which the (unknown) continuous measure is realized. For example, the latent continuous measure might be "time spent participating in activity x over time period y (say t)," but owing to whatever reasons, all one knows is whether $t=0$, $t\varepsilon(0, 4 \text{ days}]$, $t\varepsilon(4 \text{ days}, 8 \text{ days}]$, or $t\varepsilon(8 \text{ days}, 365 \text{ days})$ (for y=one year). The purpose of this section is to present an estimating technique designed to handle such situations.

The method is based on the work of Rosett and Nelson (RN) (1975), who developed what is known as the two-limit probit estimation technique, and of Stewart (1983), who generalized the RN method to account for multi-interval data. We will, therefore, refer to the model expounded here as the RNS method. We begin by positing the existence of normally-distributed random variables $Y_i^* - NID(X_i\beta, \sigma^2)$. The realizations y_i^* are unobserved, however. Only the knowledge that the realization y_i^* is an element of some proper subset of R is available. More formal1 y, partition R into $P(\geq 2)$ subsets J_k, such that $\cup J_k = R$, $J_k \cap J_j = \emptyset, \forall k,j$. The data available to the analyst are: k (such that $y_i^* \epsilon J_k$)), $\inf(J_k)$, and $\sup(J_k)$.[7] Note that when P=2 this reduces to the binary probit model while for P=3, the RN two-limit probit model emerges.

Following Stewart, we define the p-th interval by $(A_{p-1}'A_p)$, and set $A_0 = -\infty$, $A_p = +\infty$. Given T independent observations, the log-likelihood function of this model can be written

$$\ell = \sum_{i=1}^{T} \log (\phi_{p(i)} - \phi_{(p-1)(i)}) \tag{61}$$

where $\phi_{p(i)} = \phi(A_p - X_i\beta/\sigma)$, ϕ being the cumulative distribution of the standard normal. Estimation is by maximum likelihood. The first-order conditions for maximizing ℓ are

$$\partial \ell / \partial \beta = \sum_{i=1}^{T} (\phi_{(p-1)(i)} - \phi_{p(i)})/\sigma(\phi_{p(i)} - \phi_{(p-1)(i)})X_i' = 0$$

and $\tag{62}$

$$\partial \ell / \partial \sigma = \sum_{i=1}^{T} (\theta_{(p-1)(i)} - \theta_{p(i)}/\sigma(\phi_{p(i)} - \phi_{(p-1)(i)}) = 0,$$

where $\theta_{p(i)} = (A_p - X_i\beta/\sigma)(\phi(A_p - X_i\beta/\sigma))$, and $\phi(c)$ is the standard normal density $(2\pi)^{-1}\exp(-.5c^2)$. Because the matrix of second derivatives of ℓ is fairly complicated, we have elected to use for optimization purposes the method of Berndt, Hall, Hall, and Hausman (1974), which utilizes only the first derivative vector $((\partial\ell/\partial\beta)', (\partial\ell/\partial\sigma))'$. (Note that when $P = 2$, i.e. when the model is binary probit, a parameter normalization is required. Typically $\sigma=1$ is used. This reduces the number of first order conditions from (m+1) to m, where m is the dimensionality of β.) Stewart has shown how iterative least squares can be used to obtain the ML estimate. The reader is referred to his work for the details.

GROUPED-DEPENDENT VARIABLE ESTIMATION: SOME EXTENSIONS

As discussed earlier, Stewart (1983) has proposed several approaches to parameter estimation in situations where the dependent variable is grouped. These are cases where the only available information on the dependent variable is of which of P mutually exclusive and exhaustive subintervals of the real line it is an element. The main purpose of Stewart's paper is to suggest methods of consistent parameter estimation in the grouped dependent variable (GDV) model that are computationally less burdensome than are iterative maximum likelihood techniques. The intent of this section is to propose extensions of the idea of GDV estimation in several directions.

The strategy of this section is as follows. First, the GDV model is discussed in the context of the censored- and truncated-normal models for continuous dependent variables. The analogies are highlighted, and it is shown that a form of misspecification that precludes consistent parameter estimation in the censored-normal (Tobit) model might well plague GDV

estimates in sane circumstances. An interpretation of and tests for an obvious alternative specification in the spirit of Cragg's (1971) hurdles model are suggested.

Second, it is proposed that the GDV framework can, under suitable circumstances, permit consistent parameter estimation when continuous censored- or truncated-normal dependent variables are measured with error. Stapleton and Young (1984) have demonstrated that, unlike the case of the basic linear model, errors of measurement on censored or truncated dependent variables result in inconsistent parameter estimates when estimation is by maximum likelihood. It is suggested below that if the errorless dependent variable can be reasonably assumed to occupy certain intervals with probability one, consistent estimation is possible within the GDV framework: This result is particularly important in the case where the dependent variable is truncated-normal because easily-computable consistent estimators based on expectation functions (see Stapleton and Young (1984)), feasible in the censored-normal model, are more difficult to implement in the truncated case. In this context, a Hausman (1978) test is proposed for testing the errors-in-dependent-variable hypothesis.

We turn to a brief recapitulation of the basic elements of the GDV model and its estimation by means of maximum likelihood. The presentation. parallels closely that of Stewart's Section 2.

We assume the existence of T independent drawings from random variables Y_i^* ~ $N(X_i \beta, \sigma^2)$ where X_i and ε' are 1xk vectors. However, as discussed above, the point realizations are unknown to the researcher. The only information on the realizations is of which of P mutually exclusive and exhaustive subintervals of the real line it is a member. Given (P+1) constants a_p, $p \varepsilon \{0, 1, \ldots, P\}$ the P intervals are defined by I, = $(a_0, a_1]$

$\ldots,\; I_p = (a_{p-1},\; a_p),\quad$ where $a_0 = -\infty$, $a_P = +\infty$; $\quad \overset{P}{\underset{p=1}{U}} I_p = R$; $I_p \cap I_{p'} = \varnothing$, $\forall$ $p, p' \in \{1, \ldots, P\}$. For notational ease we adopt the convention $q = p-1$. Letting ϕ and Φ denote the standard normal density and distribution functions, respectively, the i-th observation's contribution to the sample likelihood is

$$L_i = \Phi((a_{p_i} - X_i \beta)/\sigma) - \Phi((a_{q_i} - X_i \beta)/\sigma) \tag{63}$$

where a_{p_i} (a_{q_i}) is the supremum (infimum) of the interval of which y_t is an element. The interpretation of (63) is L_i is the probability that a standard normal variate is in the interval $(a_{q_i},\; a_{p_i})$. The sample log-likelihood function, then, is

$$\ell = \sum_{i=1}^{T} \log \left(\Phi((a_{p_i} - X_i \beta)/\sigma) - \Phi((a_{q_i} - X_i \beta)/\sigma) \right)$$

$$= \sum_{i=1}^{T} \log (\Phi_{p_i} - \Phi_{q_i}) \tag{64}$$

The maximum likelihood estimates are obtained via simultaneous solution of the $(k+1)$ equations (65) for $(\hat{\beta}, \hat{\sigma})$:

$$\partial \ell / \partial \beta = \sum_{i=1}^{T} [(\phi_{q_i} - \phi_{p_i})/\sigma(\Phi_{p_i} - \Phi_{q_i})] X_i' = 0$$

$$\partial \ell / \partial \sigma = \sum_{i=1}^{T} ((a_{q_i} - X_i \beta) \phi_{q_i} - (a_{p_i} - X_i \beta) \phi_{p_i})/ \tag{65}$$

$$\sigma^2 (\Phi_{p_i} - \Phi_{q_i}) = 0,$$

where ϕ_{p_i} is the standard normal density evaluated at $(a_{p_i} - X_i \beta)/\sigma$ (and ϕ_{q_i} is the same density evaluated at $(a_{q_i} - X_i \beta)/\sigma$). Because the second derivatives are messy, optimization via the Berndt-Hall-Hall-Hausman method of first derivatives is an appealing choice.

The GDV model as defined above is described in a fairly general form where it is assumed that the P intervals I_p mutually exhaust the real line and that $Pr(y_i \epsilon I_p) > 0 \ \forall \ p \epsilon \{1, \ldots, P\}$. This structure can be amended, however, to allow for situations often interesting to economists. To this end, two restricted versions of the above framework, which we term the censored-GDV (CGDV) and truncated-GDV (TGDV) models, are proposed. The censored version in fact turns out to be (64), but is described independently to facilitate discussion of and comparison with the truncated specification. In both cases discussion is confined to the normally distributed case although other options are certainly available.

The "difference" between the general specification (64) and the CGDV formulation is that for the CGDV model there is a mass point of the distribution at $\sup(I_1)$ for the censored-from-below CGDV or at $\inf(I_p)$ in the censored-from-above version. That is, $Pr(y_i \epsilon I_1) = Pr(y_i = \sup(I_1))$ in the former case, $Pr(y_i \epsilon I_p) = Pr(y_i = \inf(I_p))$ in the latter case. The CGDV model can thus be viewed as a Tobit model in which the noncensored observations are grouped. The log-likelihood function is identical to (64), but can be written in a form more closely resembling the familiar Tobit log-likelihood function as

$$\ell = \sum_{i \epsilon \Omega_0} \log \left(\Phi((a_c - X_i \beta)/\sigma) \right) + \sum_{i \epsilon \Omega_1} \log \left(\Phi_{p_i} - \Phi_{q_i} \right), \tag{66}$$

where Ω_0 and Ω_1 are the index sets for the censored and noncensored observations, respectively, and where censoring from below at a_c is assumed.

Since

$$\lim_{\Delta a \to 0} \frac{(\Phi_{p_i} - \Phi_{q_i})}{\Delta a} = \frac{d\Phi(Z)}{dZ} = \phi(Z),$$ (67)

where $\Delta a = a_{p_i} - a_{q_i}$, (66) is seen to reduce to the standard Tobit model when the lengths of the intervals I_p become infinitesimally small and the number of intervals goes to infinity. When structured thusly, the CGDV model would be appropriate in situations where, for example, nonnegative data on y_i are grouped as $\{0\}$, $(1,4]$, $(4,8]$, $(8, +\infty)$, representing perhaps expenditures or hours of labor supplied. Note that the CGDV model is analytically identical to the basic GDV model (64), so that the estimation techniques suggested by Stewart can be utilized.

The TGDV model requires different treatment, however. Here, as in the continuous truncated-normal model, there is no mass point of the density like that occurring in the CGDV specification. Rather, there is assumed to be a known point of either upper or lower truncation, a_{T_i}, and we assume $Pr(y_i = a_T) = 0$. Confining discussion to truncation from below, again define P intervals I_p, now requiring $\sum_{p=1}^{P} I_p = (a_T, +\infty)$, with $a_T > -\infty$. The truncation of the density necessitates an amendment to the log-likelihood function (64), viz.

$$\ell = \sum_{i=1}^{T} \log((\Phi_{p_i} - \Phi_{q_i})/(1 - \Phi((a_T - X_i\beta)/\sigma)))$$ (68)

$$= \sum_{i=1}^{T} \log((\Phi_{p_i} - \Phi_{q_i})/\Phi((X_i\beta - a_T)/\sigma)).$$

The maximum likelihood estimates of $(\hat{\beta}, \sigma)$ satisfy

$$\partial \ell / \partial \beta = \sum_{i=1}^{T} (((\phi_{q_i} - \phi_{p_i})/\sigma(\Phi_{p_i} - \Phi_{q_i})) - (\phi_{T_i}/\sigma\Phi_{T_i})) X_t' = 0$$

$$(69)$$

$$\partial \ell / \partial \sigma = \sum_{i=1}^{T} ((a_{q_i} - X_i\beta)\phi_{q_i} - (a_{p_i} - X_i\beta)\phi_{p_i})/\sigma^2(\Phi_{p_i} - \Phi_{q_i}))$$

$$+ ((X_i\beta - a_T)\phi_{T_i}/\sigma^2\Phi_{T_i}) = 0,$$

where ϕ_{T_i} and Φ_{T_i} are ϕ and Φ evaluated at $(X_i\beta - a_T)/\sigma$. As is the case in the CGDV model, estimation based on first derivatives is a sensible approach; The TGDV model arises in situations similar to those analyzed by Stewart where, for example, earnings data are obtained for some sample in a grouped manner, but only those reporting positive earnings are sampled. The intervals might then be $(0, 10], (10,20], \ldots , (105, +\infty)$.

Having formally juxtaposed the CGDV and TGDV models, we now discuss how a form of misspecification that arises in the continuous censored-normal model can also corrupt the CGDV model. One resolution of this specification problem rests on the use of the TGDV model. Specifically, we are concerned with the grouped data analog of the case discussed by Cragg (1971) and Lin and Schmidt (1984). (With little loss of generality, we consider the censored-from-below specification in the sequel.) The central question to be addressed is whether the statistical models for the latent random variables Y_i^* determining $Pr(y_i > a_c)$ (the first-stage model) differ from the second-stage models for the conditional densities $f(y_i | y_i > a_c)$ in the continuous case or $Pr(y_i \in I_p | y_i \in I_1)$ in the GDV case. The continuous Tobit and the CGDV specifications tacitly assume that these first- and second-stage models reduce to a single model determined by the same parameterization of the mean and variance of the latent Y_i^*.

However, as discussed in depth earlier, Lin and Schmidt (1984) have presented cogent arguments against the plausibility of this tacit assumption as it pertains to many phenomena of interest to econometric modelers. Should the distribution determining the phenomenon $\Pr(y_i > a_c)$ in fact be parameterized differently than the distribution $f(y_i \mid y_i > a_c)$ or $\Pr(y_i \epsilon I_p \mid y_i \epsilon I_1)$, then maximum likelihood estimation of the standard Tobit or, of more immediate concern, the CGDV model will generally yield inconsistent estimates of the parameters of both the first- and second-stage models since the log-likelihood function is misspecified.

In the continuous case when $a_c = 0$, Cragg suggests two specifications for the second-stage conditional density of the y_t. First is a lognormal specification, where given $y_i > 0$, $\log Y_i \sim N(X_i \beta_2, \sigma^2)$. Second is a truncated-normal model, where given $y_i > 0$, $Y_i \sim TN(X_i \beta_2, \sigma^2 ; 0, +\infty)$. In both cases, estimation of the first-stage model for $\Pr(y_i > 0)$ versus $\Pr(y_i \leq 0)$ is a standard binary probit model based on some latent random variable that is distributed $N(X_i \beta_1, 1)$.

A grouped-data analog of Cragg's probit/truncated-normal model can be defined and considered as a logical alternative to the CGDV specification. Denoting β_1 and (β_2, σ^2) as the parameters of the probit first-stage and conditional truncated-normal second-stage models, respectively, the log-likelihood function of the grouped-data Cragg specification (CRGDV) is

$$\ell = \sum_{i \epsilon \Omega_0} \log(1 - \Phi(X_i \beta_1)) + \sum_{i \epsilon \Omega_1} [\log(\Phi(\frac{a_{p_i} - X_i \beta_2}{\sigma}) - \Phi(\frac{a_{q_i} - X_i \beta_2}{\sigma})) +$$

$$\log \Phi(X_i \beta_1) - \log \Phi(\frac{X_i \beta_2}{\sigma})] \tag{70}$$

$$=[\sum_{i\epsilon\Omega_0} \log(1-\Phi(X_i\beta_1)) + \sum_{i\epsilon\Omega_1} \log\Phi(X_i\beta_1)] + [\sum_{i\epsilon\Omega_1} \log(\Phi(\frac{a_{p_i}-X_i\beta_2}{\sigma})-$$

$$[\Phi(\frac{a_{q_i}-X_ib_2}{\sigma})))-\log\Phi(\frac{X_ib_2}{\sigma})] \qquad (71)$$

The maximum likelihood estimates of (β_1,β_2,σ) can be obtained by the two stage method:

    1) Probit estimation of β_1 based on the terms in the first square brackets in (71);

    2) TGDV estimation as described earlier based on the terms in the second square brackets.

Furthermore, note that when $\beta_1 = \beta_2/\sigma$, (70) reduces to the log-likelihood function of the CGDV model (66).

One can test for the appropriateness of the CRGDV model vis-a-vis the CCDV specification in several ways. First is a likelihood ratio test based on estimation of both specifications. Since the CGDV model imposes k parameter restrictions of the form $\beta_{1_j} = (\beta_{2_j}/\sigma)$, $j=1,...,k$, on the CRGDV model, the statistic $-2(\ell_{CGDV}-\ell_{CRGDV})$ is distributed as asymptotically) central χ^2 with degrees of freedom k under the null hypothesis that CGDV is the appropriate specification. $(\ell_Z$ is the maximized log-likelihood function value under specification Z). Second, along the lines suggested by Lin and Schmidt in the continuous dependent variable case, one can. design a Lagrange multiplier test for the appropriateness of the restrictions implied by the CGDV specification. Because such a test relies only on estimation of the restricted (CGDV) specification rather than of both models, it is an appealing alternative when the CGDV model is easily estimated. As the details of the derivation of this test statistic would

parallel closely the Lin and Schmidt work, they are not sketched here. Their test is in fact the limiting version of that suggested here as the size of the intervals goes to zero and their number to infinity. Finally, a Wald test analogous to that suggested earlier in the analysis of the continuous version of the Cragg model can be easily constructed.

To summarize, the TGDV model has been proposed as a useful estimation technique both when the unconditional distribution of the dependent variable is truncated and grouped in intervals and when the conditional probability distribution of grouped data is truncated, this occurring in one instance when certain restrictions implied by the GDV/CGDV specification are untenable.

We turn now to a discussion that casts GDV estimation in an entirely different role: Here focus is on situations where point data of the latent Y_i^* variates are in fact recorded. The data may be such that these realizations are of a censored or truncated nature, but in the most general of cases all points on the real line are candidates. The problem of interest here is that it is possible that the data are recorded or measured with error.

It is well known that if the independent variables are measured without error and the nature of the dependent variable is such that it is both realized over the entire real line and that measurement errors therewith associated are stochastic, additive, and have null expectation, then least squares provides consistent parameter estimates. The reason of course is that the additive measurement errors on the y_i serve only to change the variance of the additive model error, leaving unaffected all requisite conditions for consistent estimation.

However, Stapleton and Young (SY) (1984) have demonstrated that when the dependent variable is not realized over the entire real line, but is rather of a censored or truncated nature, maximum likelihood estimation yields inconsistent parameter estimates when the dependent variable is measured with error: The version of the SY model we consider is

$$Y_i^* = X_i \beta + \varepsilon_i \tag{72}$$

$$\begin{aligned} y_i &= y_i^* + v_i, \text{ rhs of (72)} > 0 \\ &= 0, \qquad \text{else} \end{aligned} \tag{73}$$

$$\begin{aligned} y_i^T &= y_i^*, \; y_i^* > 0 \\ &= 0, \qquad \text{else} \end{aligned} \tag{74}$$

$$\begin{aligned} d_i &= 1, \qquad y_i^* \leq 0 \\ &= 0, \qquad \text{else} \end{aligned} \tag{75}$$

Define the censored-SY (CSY) model as that resulting when the y_i are recorded for ail i and the truncatec-ST (TSY) model as that resulting when only the positive y_i are used. The Y_i^T are the true but unobserved positive-censored realizations of the Y_i^*. The indicators d_i are assumed available; these give information about the underlying structure of the latent errorless classification mechanism. That is, assuming all recorded measurements of y_i are nonnegative, we know when $y_i=0$ because $y_i^* \leq 0$ versus when $y_i=0$ because $y_i^*>0$ and $v_i<-y_i^*$ An example of this type of scheme is where it is known with certainty whether or not a person participated in the labor force at time t, but the number of hours of

participation--necessarily recorded as nonnegative--is possibly measured with error. Cast thusly, it is probably true that the incidence of similar phenomena in other areas of microdata analysis is significant.

SY have shown that although maximum likelihood estimation of either the censored model based on all observations or the truncated model based only on the positive y_i yields inconsistent parameter estimates, it is possible in the censored data case to obtain consistent parameter estimates via a variety of two-step techniques. In its most familiar form, associated with Heckman's work, a first-stage binary probit model on the d_i is estimated. The results from this are used to construct estimates of $\phi(X_i \beta/\sigma)$ and $\Phi(X_i \beta/\sigma)$ which in turn are used to construct the expectation function (EF) of Y_i

$$E(Y_i) = X_i \beta \Phi_i + \sigma \phi_i \qquad (76)$$

or the conditional expectation function (CEF) of y_i

$$E(y_i | d_i = 0) = X_i \beta + \sigma \phi_i / \Phi_i . \qquad (77)$$

Assuming $E(v_i | y_i^*) = 0$, SY demonstrate that least-squares estimation of (β, σ) in (76), (77), or in several other possible formulations yields consistent estimates.

When the data are truncated rather than censored, OLS estimation via the EF or CEF methods proposed by SY is no longer feasible. Consistent estimates of ϕ_i and Φ_i cannot be obtained by probit since $d_i = 0 \; \forall_i$ in the truncated case. Thus the EF and CEF cannot be estimated by OLS (see Maddala (1983, p. 167)) SY do propose a truncated nonlinear least squares

method for obtaining consistent estimates based only on positive y_i. Such a technique is highly nonlinear, however, and potentially difficult to estimate.

Given certain assumptions about the nature of the measurement errors, we now show how the parameters of both the TSY and CSY models can be estimated by maximum likelihood GDV methods. The argument is as follows. The source of the inconsistency of the maximum likelihood estimates of the TSY and CSY models is the measurement error on the y_i^*. The likelihood function formulated on the assumption that the observed y_i are measured without error is therefore based on incorrect contributions of each observation to the total likelihood when measurement errors are present, maximum likelihood thus resulting in inconsistent parameter estimates. (The appendix of SY gives a detailed proof of the inconsistency of maximum likelihood.)

Consider now the possibility that the measurement error structure is such that there exist known nonnegative scalars CL_i and CU_i such that

$$\int_{L_i}^{U_i} dG(v_i) = 1, \tag{78}$$

where $U_i = CU_i - y_i$, $L_i = CL_i - y_i$, and $G(v_i)$ is the distribution function of the measurement errors v_i. We allow for the possibility that there exist $CL_i' > CL_i$, $CU_i' < CU_i$ such that

$$\int_{L_i'}^{U_i'} dG(v_i) = 1, \tag{79}$$

$U_i' = CU_i' - y_i$, $L_i' = CL_i' - y_i$, i.e. CL_i and CU_i are not necessarily the inf and sup of the support of the density $g(v_i$. Although the requirement that CL_i

and CU_i are known is somewhat restrictive and does not admit certain forms for the density $g(v_t)$, it is essential to the following argument. To make the argument nontrivial, we assume that the intervals (CL_i, CU_i) differ across i and that there exist i,j such that $CU_i \neq CU_j$ for some i, j and $CL_i \neq CL_j$ for some (possibly other) i,j, this second requirement necessary to assure the boundedness of the likelihood function. Define $I_i = (CL_i, CU_i)$, so it is assumed that $Pr(y_i^T \epsilon I_i | y_i \epsilon I_i) = 1$. Heuristically, this means that the data and measurement error structure are such that it can be said that given y_i, y_i^T falls between CL and CU_i with certainty, i.e. the probability of "misclassifying" y_i^T is zero. The plausibility of this assumption will of course differ across empirical applications.

Given the above assumptions on the measurement error structure, it may be demonstrated that maximum likelihood estimation of either the CSY or TSY models via CGDV or TGDV techniques is a feasible approach to consistent parameter estimation when measurement errors may be present. Defining the intervals (CL_i, CU_i) analogous to (a_{q_i}, a_{p_i}) above, but no longer imposing the restriction that these intervals be established ex ante, estimation can proceed in the manner of equations (63)-(65), with now

$$L_i = \Phi((CU_i - X_i\beta)/\sigma) - \Phi((CL_i - X_i\beta)/\sigma),$$ (80)

and

$$\ell = \sum_{i=1}^{T} \log (L_i).$$ (81)

The maximum likelihood estimates derived by maximizing (81), given the assumption that the misclassification probabilities are zero, are consistent, the argument following exactly that for the CGDV and TGDV models described earlier.

There is, of course, a tradeoff of sorts involved here. The larger the intervals (CL_i, CU_i), the smaller will be the misclassification probabilities, in general. However, the larger the intervals, the less efficient will be the parameter estimates as information about the actual magnitudes of the y_i^T is lost. Note that when no measurement error is present, consistent estimates still obtain, but are not least-variance. Because the consistency properties of estimating the CSY or TS models via GDV methods rely on zero misclassification probabilities, the efficiency tradeoffs are probably worthwhile in many circumstances. This approach is particularly promising in the TSY model given that SY have demonstrated an array of computationally simple methods for estimating the CSY model but only one method for estimating the TSY model, this being potentially burdensome to estimate.

An outgrowth of the preceeding discussion, and indeed of the SY discussion as well, is that a straightforward test for the measurement error problem is available. Given that under the null hypothesis of no measurement error, maximum likelihood estimation of either the CSY or TSY models yields consistent and efficient parameter estimates; that such estimates are in general inconsistent in the presence of measurement error; and that regardless of the presence of measurement error, there exist estimators that are consistent but inefficient, then a Hausman (1978) test is suggested. These consistent-inefficient estimators, of course, are the EF or CEF estimators suggested by SY and, given the appropriate assumptions on the measurement error structure, the CGDV/TGDV estimators suggested above. The form of the Hausman test is

$$m = T(\hat{\beta}_1 - \hat{\beta}_0)'(\hat{V}_1 - \hat{V}_0)^{-1}(\hat{\beta}_1 - \hat{\beta}_0), \tag{82}$$

where under H_0: no measurement error, $\hat{\beta}_1$ is the consistent-inefficient estimate of β, $\hat{\beta}_0$ is the efficient maximum likelihood estimate, and $\hat{V}_1$ and $\hat{V}_0$ are the corresponding estimates of the covariance matrices. When the EF or CEF estimators are used for $\hat{\beta}_1$, the appropriate formulae for $\hat{V}_1$ can be found in SY; when the CGDV or TGDV approach is taken, the appropriate submatrix of the inverse of

$$\sum_{i=1}^{T} (\partial\log L_i / \partial\theta)(\partial\log L_i / \partial\theta'), \tag{83}$$

can be used to estimate $\hat{V}_1$, where $\theta = (\beta', \sigma)'$, L_i is as defined in (80), and evaluation of θ is at the CGDV or TGDV estimates. Under the null hypothesis of no measurement error, m is distributed asymptotically central χ^2 with k degrees of freedom.

Furthermore, note that the version of the Hausman test proposed by Nelson (1981), whose extension to the truncated normal case was discussed above, can also be used as a test of the measurement error hypothesis. Since both the censored and truncated versions of the Nelson test rely on estimates of $E(X'Y)$ to form the test statistic, and since Nelson's consistent-inefficient moment estimator of this expectation remains consistent given null expectation of the measurement errors, then his test qualifies as a Hausman test for the measurement error problem. However, because the Nelson tests are appropriate tests for a wide variety of problems (heteroscedasticity, nonnormality, to name two), a significant Nelson statistic will not necessarily shed light on the nature of the problem it has diagnosed.

PREDICTION IN ESTIMATED PARTICIPATION MODELS

The preceeding sections have surveyed some econometric methods for estimating recreational participation models. Because there exists a variety of structural or behavioral assumptions about the mechanisms that give rise to the statistical formulations, as well as a variety of data collection methods and coding configurations, it has been necessary to consider a set of possible approaches to estimation strategy. The models considered are largely of a non-nested nature (i.e. Model A cannot generally be obtained as a restricted version of Model B and vice-versa). And the techniques for non-nested model evaluation are largely undeveloped in situations where the model error structures are nonnormal. Thus, it is necessary that comparisons be made in terms of alternative predictions across specifications if policy is to be guided sensibly.

While the goal in much of this chapter has been the goal to obtain consistent parameter estimates of participation models, specified, in and of themselves, consistently estimated models are nothing more than aesthetically-pleasing curiosa. Their raison d'être insofar as the present analysis is concerned is of course to serve as tools for predicting the impacts of changes in water quality or recreation participation.

In the practical realm of policy analysis, Intriligator (1983) refers to such prediction methods as the simulation approach to policy evaluation and summarizes the approach as follows:

> This approach uses the estimated reduced form to determine alternative combinations of policy variables and endogenous variables for a given set of possible policies...
> The policymaker would provide the model builder with the alternative policies, and the model builder would, in turn,

provide the decision maker with their consequences for the endogenous variables. The policymaker would then choose a desired policy and its outcome...

This approach requires that the policymaker formulate an explicit set of policy alternatives and that an estimated econometric model incorporating the appropriate policy variables be available. Simulation, based in part on communication between policymaker and model builder, represents a valuable approach to policy evaluation that could be used in any policy area in which there exists a relevant estimated econometric model. (Intriligator (1983), p. 214).

We think it necessary to close this chapter with a discussion of prediction in the context of the econometric models discussed earlier because the relevant prediction formulae and methods vary considerably across the model specifications. Though the results presented below are hardly profound, their presentation merits the space used basically because there exists to our knowledge no unified treatment of prediction that includes the several econometric models proposed above. Such a unified treatment should be of interest to the policy analyst interested in juxtaposing the estimated policy outcomes from the various econometric specifications, which are, as noted above, largely nonnested.

Prediction in the context of the econometric participation models discussed in this chapter is the process whereby one assesses the change in the estimated response with respect to a change in some control variable, specifically water quality. The statistical models estimated are typically of a nature for which it is possible to describe one or more of the following:

a) the expectation of the dependent variable, $E(Y_i)=f(X_i\beta)$;

b) the conditional expectation of the dependent variable, e.g.,

$E(Y_i|y_i>0)=g(X_i\beta)$;

c) the probability that the dependent variable equals sane value, e.g., $Pr(Y_i=y_i)=h(X_i\beta)$ (this description pertains mainly to qualitative dependent variable models such as the multinomial model described earlier).

For example, the objective of the econometric estimation of some model might be to obtain an estimate of $E(Y_i)$ as $\hat{Y}_i=f(X_i\hat{\beta})$. Prediction, then, would be the process whereby one estimates $\partial E(Y_i)/\partial x_{ik}$ by $\partial \hat{Y}_i/\partial x_{ik}$, with X_{ik} some control variable specific to i. In the subsequent discussion, we take the term "prediction" generally to mean the expected change in y attributable to a hypothetical change in sane control variable X, i.e., $\partial E(y)/\partial X$. Frequently, the estimated changes are couched in terms of elasticities in order to abstract from magnitudinal considerations; here one might use estimates of $\partial \log E(Y_i)/\partial \log x_{ik}$. We turn now to an analysis of how prediction of the above nature would apply to the econometric models discussed in this chapter.

The ordinary least squares (OLS) specification is an obvious starting point. Although we have seen how OLS will generally be an inappropriate analytical tool given the nature of most participation data likely to be encountered, the analytics of prediction in the OLS model are quite simple and serve to motivate the remainder of the discussion.

Recall that the basic linear model can be written as

$$Y_i = X_i\beta + \varepsilon_i \tag{84}$$

If we have $E(\varepsilon_i)=0$, then $E(Y_i)=X_i\beta$. OLS estimates β by $\hat{\beta} = (X'X)^{-1}X'y$: estimates $E(Y_i)$ by $\hat{Y}_i=X_i\hat{\beta}$. Note that $E(\hat{Y}_i)=E(X_i\hat{\beta})=X_i(X'X)^{-1}X'(X\beta+E(\varepsilon))$:

$=E(Y_i)$ so that $\hat{Y}_i$ in the OLS model is an unbiased estimator of $E(Y_i)$ given the ideal conditions described earlier in this chapter. Prediction of the expected change in y_i for a change in X_{ik} in this case is simply $\partial \hat{Y}_i / \partial x_{ik} = \hat{\beta}_k$, or the k-th slope estimate. In the linear model, it is fairly standard practice to use the elasticity $\partial \log Y_i / \partial \log x_{ik}$ and evaluate the elasticity at the means of the observed y_i and x_{ik}, viz. $\hat{\beta}_k \bar{X}_k / \bar{Y}$, where the overbar denotes a sample mean.

The Tobit model lends itself to description by either its expectation function or its conditional expectation function. Respectively, these are

$$E(Y_i) = X_i \beta \Phi_i + \sigma \phi_i \qquad (85)$$

and

$$E(Y_i | y_i > 0) = X_i \beta + \sigma \phi / \Phi_i, \qquad (86)$$

where ϕ_i and Φ_i are evaluated at $X_i \beta / \sigma$. Following Maddala (1983), the respective predictions are

$$\partial E(Y_i) / \partial x_{ik} = \Phi_i \beta_k \qquad (87)$$

and

$$\partial E(Y_i | y_i > 0) / \partial x_{ik} = \beta_k (1 - (X_i \beta \phi_i / \Phi_i \sigma) - (\phi_i / \Phi_i)^2) \qquad (88)$$

In elasticity form, these partial derivatives are multiplied by $x_{ik}/E(Y_1)$, and $x_{ik}/E(Y_i | y_i > 0))$, respectively. Evaluation, again, might be at the sample means of the y_i and X_i. However, note that in this specification, as in other nonlinear-in-parameters specifications, the evaluated predictions will typically depend not only on the selected y and x_k evaluation points, but also on the other (noncontrol) elements of the X-vectors. In the general nonlinear case, evaluation at the means of the X_i does not generate the same prediction as is yielded by calculating the mean of the estimated individual predictions. This is merely a corollary of a general property of nonlinear functions where $E(f(Z)) \neq f(E(Z))$.

Thus, another obvious method of evaluation would be to calculate the prediction for each individual in the sample and average the individual predictions. Such a strategy might be valuable when contemplated policy measures take the form of, say, an x-percent change in the control variable of each individual from that individual's prevailing or pre-policy level.

In the probit/truncated-normal version of the Cragg estimator (the Cragg equations (7) and (9)) the expectation function can be written as

$$E(Y_i) = E(Y_i|y_i>0)Pr(y_i>0) \tag{89}$$

$$=[X_i\beta_2 + \sigma_2\phi_{i2}/\Phi_{i2}]\Phi_{i1} \tag{90}$$

and

$$E(Y_i|y_i>0) = X_i\beta_2 + \sigma_2\phi_{i2}/\Phi_{i2},$$

where ϕ_{ip} and Φ_{ip} are the standard normal density and distribution functions evaluated at $X_i\beta_p/\sigma_p$ (p=1,2). (Recall that σ_1 is normalized = 1). In this specification,. the formula for $\partial E(Y_i|y_i>0)/\partial x_{ik}$ is identical to that for the Tobit specification, equation (86) above. The formula for the unconditional version is

$$\partial E(Y_i)/\partial x_{ik} = \beta_{2k}\Phi_{i1} + X_i\beta_2\phi_{i1}\beta_{1k} + \sigma_2\lambda_{i2}[\phi_{i1}\beta_{1k} -$$

$$\tag{91}$$

$$(\phi_{i1}Z_{i2}\beta_{2k}/\sigma_2) - (\lambda_{i2}\phi_{i1}\beta_{2k}/\sigma_2)],$$

where $\lambda_{i2} = \phi_{i2}/\Phi_{i2}$ and $Z_{i2}= (X_i\beta_2/\sigma_2)$. Note that when $\beta_1=\beta_2$ and $\sigma_1 = \sigma_2$, i.e., when the Tobit restrictions are imposed on this version of the Cragg model, (91) reduces to the far less-complicated expression (87).

The second version of the Cragg model considered earlier is that where the second-stage conditional density is lognormal, with log $Y_i \sim N(X_i\beta_2, \sigma_2^2)$ conditional on $y_i>0$. Again defining the appropriate expectation functions yields

$$E(Y_i) = E(Y_i | y_i > 0) \Pr(y_i > 0)$$

(92)

$$= \exp(X_i \beta_2 + .5\sigma_2^2) \Phi_{i1}$$

and

$$E(Y_i | y_i > 0) = \exp(X_i \beta_2 + .5\sigma_2^2),$$

(93)

where we have used the well-known property of lognormal densities that where the log of a lognormally-distributed random variable has mean μ and variance θ^2, then the expected value of the random variable is $\exp(\mu + .5\theta^2)$ (See Johnson and Kotz (1970, Chapter 14) for further details on the properties of the lognormal). The relevant partial derivatives are

$$\partial E(Y_i)/\partial x_{ik} = \exp(X_i \beta_2 + .5\sigma_2^2) [\phi_{i1} \beta_{1k} + \Phi_{i1} \beta_{2k}]$$

(94)

and

$$\partial E(Y_i | y_i > 0)/\partial x_{ik} = \exp(X_i \beta_2 + .5\sigma_2^2) \beta_{2k}.$$

(95)

In all cases, elasticity forms are calculated by the appropriate multiplication by either $x_{ik}(E(Y_i))^{-1}$ or $x_{ik}(E(Y_i | y_i > 0))^{-1}$ of the partial derivative formulae derived above.

The Heckman two-equation model, recall, is cast in terms of unconditional densities of the random variables Y_{i1} and Y_{12}^*. Thus, tne unconditional expectation of Y_{i1} is a linear function of parameters,

$$E(Y_{i1}) = X_i \beta_1,$$

(96)

with

$$\partial E(Y_{i1})/\partial x_{ik} = \beta_{1k}$$

(97)

If the moment of interest is the conditional expectation of Y_{i1} given $y_{i2} = 1$ (in the notation of equation (34)), then we have from equations (35) and (36)

$$E(Y_{i1} | y_{i2} = 1) = X_i \beta_1 + \sigma_{12} \phi_{i2}/\sigma_2 (1 - \Phi_{i2}),$$

(98)

where ϕ_{i2} and Φ_{i2} are the standard normal density and distribution

functions evaluated at $(X_i\beta_2/\sigma_2)$. Here,

$$\partial E(Y_{i1}|y_{i2}=1)/\partial x_{ik} = \beta_{1k} + (\sigma_{12}\lambda_{i2}\beta_{2k}/\sigma_2^2)[\lambda_{i2}-Z_{i2}], \qquad (99)$$

where now we define $\lambda_{i2} = \phi_{i2}/(1-\phi_{i2})$ and $Z_{i2} = X_i\beta_2/\sigma_2$.

The derivations of the prediction formulae for the Poisson model are straightforward. Recall from equation (49) that the expectation function for the Poisson specification is

$$E(Y_i) = \exp(X_i\beta). \qquad (100)$$

(Note that the Poisson specification differs from the others discussed above in that the conditional expectation function has not been considered. [8])

Thus we have

$$\partial E(Y_i)/\partial x_{ik} = \beta_k\exp(X_i\beta) \qquad (101)$$

or, in elasticity form,

$$\partial E(Y_i)x_{ik}/\partial x_{ik}E(Y_i) = \beta_k X_{ik}. \qquad (102)$$

The formulae for the geometric specification, with $E(Y_i)=\exp(X_i\beta)$, are also (101) and (102).

Recall for the Stewart or grouped-dependent variable model that in any interval (a_{p-1},a_p) one has for a conditional expectation

$$E(Y_i|y_i\epsilon(a_{p-1},a_p)) = X_i\beta + \sigma[\phi(Z_{p-1})-\phi(Z_p)]/[\Phi(Z_p)-\Phi(Z_{p-1})], \qquad (103)$$

where $Z_p=(a_p-X_i\beta)/\sigma$. The probability that the random variable Y_i is realized in (a_{p-1},a_p) is of course $(\Phi(Z_p)-\Phi(Z_{p-1}))$. Therefore, the unconditional expectation function is

$$E(Y_i) = \sum_{p=1}^{P} E(Y_i|y_i\epsilon(a_{p-1},a_p))Pr(y_i\epsilon(a_{p-1},a_p))$$

$$(104)$$

$$= \sum_{p=1}^{P} X_i\beta(\Phi(Z_p)-\Phi(Z_{p-1})) + \sigma(\phi(Z_{p-1})-\phi(Z_p)).$$

The strong resemblance between (104) and (85) is clearly more than coincidence, such of course owing to the foundations of both the Tobit and the GDV estimators in normally-distributed latent variables.

Because the calculations are a bit messy, the prediction formulae are derived only for the expectation function; the conditional expectation variant is analogously derived. We have from (104)

$$\partial E(Y_i)/\partial x_{ik} = \sum_{P=1}^{P} [(\phi_p - \phi_{p-1}) + 2X_i \beta (\phi_p - \phi_{p-1})]\beta_k, \tag{105}$$

where $\phi_p = \phi(Z_p)$ and $\phi_p = \phi(z_p)$. The elasticity version is again derived by the appropriate multiplication.

In all the above formulations, the dependent variables or their latent bases were of a quantitative nature, thus allowing direct quantitative representation of the moments or conditional moments of interest. The prediction strategy when the dependent variable is qualitative rather than quantitative is somewhat different. However, in the one qualitative dependent variable model discussed in this chapter--the multinomial--the nature of the dependent variable is such that a fairly direct translation from the qualitative outcome to a quantitative prediction is possible. The strategy is as follows. Recall that the qualitative outcome measures of interest in the multinomial model are the probabilities of engaging in any of a set of activities on a given day. To translate these probabilities into quantitative participation measures for some time interval (a year, e.g.,), one simply sums the activity-specific day probabilities over the year. If the day participation probabilities are

Fish	.02
Boat	.01
Swim	.05
Nonrecreation	.92,

Then the annual quantitative participation measures are

Fish	.02(365) = 7.12 days
Boat	.01(365) = 3.56 days
Swim	.05(365) = 17.80 days
Nonrecreation	.92(365) = 327.52 days.

The prediction strategy, therefore, is to assess the responsiveness of the daily activity-specific probabilities to changes in the control variable, assess the ex post magnitude of annual participation in the various activities, and compare these post-policy magnitudes to those that were estimated to prevail in the pre-policy period.

The prediction equations are based on the probabilities defined in (57). For individual i and choice Z, one has

$$P_{Z_i} = \exp(X_i \beta_Z) / \sum_{j \epsilon \Omega} \exp(X_i \beta_j), \tag{106}$$

where Ω is the choice index set.

Therefore,

$$\partial P_{Z_i} / \partial x_{ik} = P_{Z_i} (\beta_{Z_k} - [\sum_{j \epsilon \Omega} \exp(X_i \beta_j) \beta_{jk} / \sum_{j \epsilon \Omega} \exp(X_i \beta_j)]). \tag{107}$$

The elasticity formulation is

$$(\partial P_{Z_i} / \partial x_{ik})(x_{ik} / P_{Z_i}) = x_{ik} (\beta_{Zk} - [\sum_{j \epsilon \Omega} \exp(X_i \beta_j) \beta_{jk} / \sum_{j \epsilon \Omega} \exp(X_i \beta_j)]). \tag{108}$$

CONCLUDING REMARKS

This chapter has discussed a variety of approaches to econometric estimation of recreation participation models. The menu of estimators

discussed while broad, hardly exhausts the set of candidates that are available. The intention here has been to be suggestive of general approaches to estimation that might be considered given heterogeneous participation data structures.

It must be stressed that because these methods generally require iterative solution algorithms, they are expensive to implement. Moreover, as the subsequent chapters show, the quality of the participation and water quality data on hand is suspect. Therefore, we decided that the potential statistical enhancements attributable to the use of these more sophisticated estimation techniques would be lost in the noise and, as such, probably not worth their added costs. Accordingly, in the empirical work to follow, more main stream techniques are utilized.

NOTES

1.   The derivation of (2) is straightforward.   Given $X \sim N(\mu, \sigma^2)$, then we wish to show:

$$E(X \mid x>0) = \mu + \sigma[f(\mu/\sigma)/F(\mu/\sigma)/F(\mu/\sigma)]$$

$$E(X \mid x>0) = \frac{1}{F} \cdot \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty x e^{-(x-\mu)^2/2\sigma^2} dx \qquad (F=F(\mu/\sigma); f=f(\mu/\sigma))$$

$$= \frac{1}{k} \int_0^\infty x e^{-(x-\mu)^2/2\sigma^2} dx \qquad (K = \sigma F \sqrt{2\pi})$$

Let $z = x - \mu$; thus

$$E(X \mid x>0) = \frac{1}{k} \{ \int_{-\mu}^\infty (z+\mu) e^{-z^2/2\sigma^2} dz \}$$

$$= \frac{1}{k} \{ \int_{-\mu}^\infty z e^{-z^2/2\sigma^2} dz + \mu \int_{-\mu}^\infty e^{-z^2/2\sigma^2} dz \}$$

$$= \frac{1}{k} \{ -\sigma^2 e^{-z^2/2\sigma^2} \Big|_{-\mu}^\infty + (\sqrt{2\pi}\sigma F)\mu \}$$

$$= \frac{\sigma^2}{k} e^{-\mu_2/2\sigma_2} + \mu$$

$$= \sigma f/F + \mu.$$

Setting $x = y_i^*$ and $\mu = x_i \beta$ yields (2).

2.   Recall that the binomal print model is the common designation models of binary (0,1) outcomes that are generated by $N(X_t \alpha, 1)$ variates. See Maddala   (1983,   chapter 2) for additional discussion.

3.   The Duan, et. al. model likelihood function (their eq. 3.7) is incorrect.   They omit the multiplicative term $(1/y_{i2})$ in the density for the positive terms, this as mentioned above being the Jacobian of the transformation from y_{i2}).   Cragg correctly incorporates this in his eq. (11).   The values of the parameters that maximize the log-likelihood function do not depend on this transform, although the value of the

log-likelihood function itself does, of course, depend on the transform.

4.   In the Lagrange-multiplier hypothesis testing approach, one formulates the problem by considering the maximization of the log-likelihood function $\ell(\theta)$ subject to a set of (perhaps nonlinear) restrictions on θ of the form $h_j(\theta)=0$.   Thus, one can consider the Lagrangean function

$$Q(\theta) = \ell(\theta) + \sum_j h_j(\theta),$$

and maximize $Q(\theta)$ w.r.t. θ.   The test relies on the idea that when the restricted estimate $\tilde{\theta}$ is "near" the unrestricted ML estimate $\hat{\theta}$, the vector $\ell(\tilde{\theta})$ will approach the zero vector.   Further discussion is found in the excellent piece by Breusch and Pagan (1980).

5.   The resemblance is due to the fact that, given the $P_i = \exp(X_i\beta)$ parameterization, if the geometric specification is reduced to the binary outcomes $\Pr(k_i=0)$ versus $\Pr(k_i = 1$ or $k_i = 1$ or...), then the binary logit model results.   This result is interesting in that consistent estimates of β in the geometric model as specified can be obtained via a binary logit model.   Such estimates are inefficient, however, as information on the magnitudes of the $k_i \gtrsim 1$ is discarded.

6.   The only difference, as is obvious from inspection of (4), is that in the one-trial case, $\Sigma t_{j_m} = 1$ for all m while in the multiple-trial case considered here $\Sigma t_{j_m} = T$.   Existing programs, then, are modified as to the number of times   the $\log P_{j_m}$ terms   are   summed in   computing   the log-likelihood.

7.   Inf   and   sup   are   the   abbreviations   for   infimum   and   supremum, respectively the greatest lower bound and least upper bound of a set S if these bounds exist.   If S has a minimum (resp. maximum) element, then $\inf(S) = \min(S)$ (resp. $\sup(S) = \max(S)$.

8. It would be possible to define a truncated version of the Poisson model were only strictly positive integer realizations were admitted. Here one would have

$$Pr(Y_i = y) = \frac{\exp(-\lambda_i)\lambda_i^y}{y!(1-\exp(-\lambda_i))}, \; y = 1,2,3,\ldots$$
$$= 0 \qquad , \text{ else,}$$

where $(1-\exp(-\lambda_i))$ is the truncation normalizing constant equal to $(1-Pr(Y_i=0))$. This could, then, be considered a conditional distribution for the $y_i \neq 0$. Some of the same questions as arise in the Lin-Schmidt critique of the Tobit model are present in the discussion of the truncated Poisson formulation. specifically, one might question whether the Poisson binary probabilities $Pr(Y_i=0)$ and $Pr(Y_i=1$ or $Y_i=2$ or $Y_i=3$ or...) $(=1-Pr(Y_i=0))$ are governed by the same statistical process as are the conditional probabilities $Pr(Y_i=y|y=1$ or $y=2$ or...). Because such considerations require considerable further development, we postpone their discussion for future research.

REFERENCES

Amemiya, T. 1981. "Qualitative Response Models: A Survey," Journal of Economic Literature, vol. 19, pp. 1483-1536.

_____. 1983. "Nonlinear Regression Models," in Z. Griliches and M. D. Intriligator, eds., Handbook of Econometrics, vol. 1, (Amsterdam: North-Holland).

_____. 1984. "Tobit Models: A Survey," Journal of Econometrics, vol. 24, pp. 3-61.

Berndt, E. R., B. H. Hall, R. E. Hall, and J. A. Hausman. 1974. "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement, vol. 3, pp. 653-665.

Breusch, T. and A. R. Pagan. 1980. "The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics," Review of Economic Studies, vol. 47, pp. 239-253.

Cox, D. R. and D. v. Hinkley. 1974. Theoretical Statistics (London: Chapman and Hall).

Cragg, J. G. 1971. "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods," Econometrica, vol. 39, pp. 829-844.

Duan, N., W. G. Manning, C. N. Morris and J. P. Newhouse. 1983. "A Comparison of Alternative Models for the Demand for Medical Care," Journal of Business and Economic Statistics, vol. 1, pp. 115-126.

_____, _____, _____, and _____. 1984. "Choosing Between the Sample-Selection Model and the Multi-part Model," Journal of Business and Economic Statistics, vol. 2, pp. 283-289.

Dudley, L. and C. Montmarquette. 1976. "A Model of the Supply of Bilateral Foreign Aid," American Economic Review, vol. 66, pp. 132-142.

Hausman, J. A. 1978. "Specification Tests in Econometrics," Econometrica, vol. 46, pp. 1251-1271.

_____, B. Hall and Z. Griliches. 1984. "Econometric Methods for Count Data with an Application to the Patents-R&D Relationship," Econometrica, vol. 52, pp. 909-938.

_____, B. Ostro and D. Wise. 1984. "Air Pollution and Lost Work," NBER working paper 1263, January.

Hay, J. W. and R. J. Olsen. 1984. "Let Them Eat Cake: A Note on Comparing Alternative Models of the Demand for Medical Care," Journal of Business and Economic Statistics, vol. 2, pp. 279-282.

Heckman, J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," <u>Annals of Economic and Social Measurement</u>, vol. 5, pp. 475-492.

_____. 1979. "Sample Selection Bias as a Specification Error," <u>Econometrica</u>, vol. 47, pp. 153-161.

Hurd, M. 1979. "Estimation in Truncated Samples When There is Heteroscedasticity," <u>Journal of Econometrics</u>, vol. 11, pp. 247-258.

Johnson, N. L. and S. Kotz. 1969. <u>Distributions in Statistics: Discrete Distributions</u> (New York: Wiley).

_____ and _____. 1970. <u>Distributions in Statistics: Continuous Univariate Distributions - I</u> (New York: Wiley).

Kendall, M. G. and A. Stuart. 1973. <u>Advanced Theory of Statistics</u>, vol. 3 (London: Griffin).

Killingsworth, M. R. 1983. <u>Labor Supply</u> (Cambridge: Cambridge University Press).

Lin, T.-F. and P. Schmidt. 1984. "A Test of the Tobit specification Against an Alternative Suggested by Cragg," <u>Review of Economics and Statistics</u>, vol. 66, pp. 174-177.

Maddala, G. S. 1977: <u>Econometrics</u> (New York: McGraw-Hill).

_____. 1983. <u>Limited-Dependent and Qualitative Variables in Econometrics</u> (Cambridge: Cambridge University Press).

Manski, C. F. and D. McFadden. 1981. <u>Structural Analysis of Discrete Data with Econometric Applications</u> (Cambridge, Mass: MIT Press).

Morey, E. R. 1981. "The Demand for Site-Specific Recreational Activities: A characteristics Approach," <u>Journal of Environmental Economics and Management</u>, vol. 8, pp. 345-371.

Nelson, F. D. 1981. "A Test for Misspecification in the Censored-Normal Model," <u>Econometrica</u>, vol. 49, pp. 1317-1329.

Olsen, R. 1980. "Approximating a Truncated Normal Regression with the Method of Moments," <u>Econometrica</u>, vol. 48, pp. 1099-1106.

Ostro, B. 1983. "The Effects of Air Pollution on Work Loss and Morbidity," <u>Journal of Environmental Economics and Management</u>, vol. 10, pp. 371-382.

Pearson, K. and A. Lee. 1908. "Generalized Probable Error in Multiple Normal Correlations," <u>Biometrika</u>, vol. 6, pp. 59-68.

Pitt, M. 1983. "Food Preferences and Nutrition in Rural Bangladesh," <u>Review of Economics and Statistics</u>, vol. 65, pp. 105-114.

Portney, P. R. and J. Mullahy. 1984. "Urban Air Quality and Acute Respiratory Illness," Resources for the Future Working Paper, September.

Rao, C. R. 1965. Linear Statistical Inference and Its Applications, (New York: Wiley).

Rosett, R. N. and F. D. Nelson. 1975. "Estimation of a Two-Limit Probit Regression Model," Econometrica, vol. 43, pp. 141-146.

Royall, A. 1984. "Robust Inference Using Maximum Likelihood Estimators," Johns Hopkins University, Department of Biostatistics Working Paper.

Schmidt, P. 1976. Econometrics (New York: Marcel Dekker).

Smith, M. and G. Maddala. 1983. "Multiple Model Testing for Non-Nested Heteroscedastic Censored Regression Models," Journal of Econometrics, vol. 21, pp. 71-81.

Stapleton, D. and D. Young. 1984. "Censored Normal Regression with Measurement Error on the Dependent Variable," Econometrica, vol. 52, pp. 737-760.

Stewart, M. B. 1983. On Least Squares Estimation when the Dependent Variable is Grouped," Review of Economics Studies, vol. 50, pp. 737-753.

Tobin, J. 1957. "Estimation of Relationships for Limited Dependent Variables," Econometrica, vol. 26, pp. 24-36.

Vaughan, William J. and Clifford S. Russell. 1982. Freshwater Recreational Fishing, The National Benefits of Water Pollution Control. Resources for the Future, Washington, D.C.

Wales, T. and A. Woodland. 1983. "Estimation of Consumer Demand Systems with Binding Non-Negativitay Constraints," Journal of Econometrics, vol. 21, pp. 263-285.

White, H. 1982. "Maximum Likelihood Estimation of Misspecified Models," Econometrica, vol. 50, pp. 1-25.

White, H. 1983. "Corrigendum," Econometrica, vol. 51, p. 513.